CUK-H06222-33-P26/296 NOV

Volume 57, Parts 1 and 2 *May 1966*

C.U.

THE BRITISH JOURNAL OF PSYCHOLOGY

CALCUTTA

EDITED BY

ARTHUR SUMMERFIELD

33

105
.C03

CAMBRIDGE UNIVERSITY PRESS
BENTLEY HOUSE, 200 EUSTON ROAD, LONDON, N.W.1
AMERICAN BRANCH:
32 EAST 57TH STREET, NEW YORK, N.Y. 10022

II

60s. net $11.00

Assistant Editors

JAMES HOGG PETER MITTLER

Preparation of Papers for the British Journal of Psychology

In submitting papers for publication in the *British Journal of Psychology* contributors are asked to observe the following requirements:

1. Articles must be as short as adequate presentation allows. It is rarely possible to publish papers more than 7000 words in length, and most papers suitable for publication in the *Journal* gain by being much shorter than this. Supplementary material of special interest may be lodged with the Technical Information and Documentary Unit, Department of Scientific and Industrial Research.
It may be possible occasionally to give early publication to very short articles, not exceeding two printed pages in length (about 1000 words). As a rule such articles need not have a summary or titled headings.

2. All papers and reviews must be typed on one side of the paper only with double spacing including list of references, and all the pages numbered. Footnotes should be avoided. Articles should where possible be sub-divided with titled headings, following the pattern of articles published from Vol. 56 (1965) onwards.
A brief summary, suitable for use as an abstract, should precede the article.

3. The Editor cannot consider for publication articles seriously deficient in presentation. It is essential that a typescript submitted should represent the final form in which the author wishes the paper to appear. Authors may be charged for any proof changes (other than printers' errors) that may involve re-setting of lines or paragraphs. Insertion of new material involving re-paging, and changes in blocks made from diagrams or drawings, are so costly that authors are well advised to regard them as out of the question.

4. Tables, photographs and figures which are costly to print should be used only when they save considerable space or are essential for effective presentation; and in no case should both a figure and table be used to convey the same information. Each table or figure should be numbered and on a separate sheet with its approximate position indicated in the typescript.
Drawings should be in Indian ink on heavy unruled paper or on graph paper ruled in light blue. Lettering on graphs, etc., is best indicated in a tracing placed over the diagram. Diagrams are normally reduced in reproduction and allowance must be made for corresponding reduction in the thickness of lines and the spaces between them. The numbered legend of a figure should be typewritten on a separate sheet of paper, not lettered on the drawing.
The object of the caption for a table or the legend for a figure should be to make the table or the figure comprehensible without reference to the text.
Authors must ensure that formulae are clear, particularly when these contain subscripts.

5. References in the text should be indicated by dates in parentheses after the authors' names and corresponding to an alphabetical list at the end of the paper. This list must give authors' names with all the details and the layout illustrated in the following examples:

ALLPORT, G. W. & VERNON, P. E. (1931). A test for personal values. *J. abnorm. soc. Psychol.* **26**, 231–48.

GALTON, F. (1883). *Inquries into Human Faculty and its Development.* London: Macmillan.

RÉVÉSZ, G. (1951). Le problème du génie. *L'Année psychol.* **50**, 83–96.

World list abbreviations are used for journals.
The Oxford Dictionary is taken as the authority on spelling.
For other general guidance, authors are referred to *The Journal of Physiology: Suggestions to Authors.* Cambridge University Press. Price 2*s*.

The British Psychological Society subscribes to the Fair Copying Declaration of the Royal Society under the terms of which reproduction of material published in this *Journal* is permitted.

THE BRITISH JOURNAL OF
PSYCHOLOGY

EDITED BY
ARTHUR SUMMERFIELD

ASSISTANT EDITORS
KEVIN CONNOLLY   PETER MITTLER

VOLUME 57, 1966

CAMBRIDGE
AT THE UNIVERSITY PRESS
1966

PUBLISHED BY
THE SYNDICS OF THE CAMBRIDGE UNIVERSITY PRESS
Bentley House, 200 Euston Road, London, N.W. 1
American Branch: 32 East 57th Street, New York, N.Y. 10022

261296

Printed in Great Britain at the University Printing House, Cambridge

CONTENTS OF VOLUME 57

PARTS 1 AND 2. MAY 1966

PAGE

Effects of intermodal stimulation on figural after-effects. By K. P. SATINDER . . . 1

Uncertainty and transitional probability in the span of apprehension. By E. K. WARRINGTON, M. KINSBOURNE and M. JAMES 7

Guiding Deutsch's model in reverse. By D. PRICHATT and D. H. HOLDING . 17

Sutherland's two-stage theory of discrimination learning and its experimental support. By W. A. MATTHEWS 25

Positive and negative recency in two-choice guessing. By H. C. A. DALE . 35

Varying the number of alternatives in short-term recall. By M. M. WOODHEAD 45

Working backward and working forward in problem-solving. By W. S. ANTHONY 53

The inferential basis of concept attainment. By N. E. WETHERICK . . 61

The relation between breadth of category and decision time. By H. TAJFEL and J. S. BRUNER 71

Memory and distortion of meaningful written material. By R. M. DAWES . 77

Relation between semantic satiation and verbal conditioning. By J. P. DAS . 87

The effect of short periods of food-deprivation on human performance. By R. A. KENNEDY and A. G. KEENE 93

The effects of distraction on pursuit rotor learning, performance and reminiscence. By H. J. EYSENCK and W. THOMPSON 99

The effect of drive on performance and reminiscence in a complex tracing task. H. J. EYSENCK and R. A. WILLETT 107

Age and perceptual-motor transfer in imbeciles: task complexity as a variable. By A. D. B. CLARKE and G. M. COOPER 113

Width of transfer and task complexity in the conceptual learning of imbeciles. By A. M. CLARKE, G. M. COOPER and A. S. HENNEY 121

Creativity and intelligence: a partial replication with Scottish children of Getzels' and Jackson's study. By P. HASAN and H. J. BUTCHER . . 129

The genetic determination of differences in intelligence: a study of monozygotic twins reared together and apart. By C. BURT 137

Personality and judgements of temporal intervals. By C. B. BELL and A. N. WATTS 155

Social desirability in individual performance on thirteen MMPI scales. By E. E. BOE and W. S. KOGAN 161

PAGE
The Brook reaction test of interests. By A. W. HEIM and K. P. WATTS . 171
Prediction of grid scores. By J. M. M. MAIR 187
Geometric illusions and environment: a study in Ghana. By G. JAHODA . 193
Induction of stereoscopic depth effects. By N. PASTORE and M. TERWILLIGER 201
The contradiction of epiphenomenalism. By E. A. LOCKE 203
PUBLICATIONS REVIEWED 205

PARTS 3 AND 4. NOVEMBER 1966

The alternation process in binocular rivalry. By W. J. M. LEVELT . . 225
EEG correlates of threshold regulation as a function of stimulus wavelength: a comparison between normal subjects and psychiatric patients. By N. F. DIXON 239
Two types of set and the generalization of perceptual defence. By A. TAYLOR and D. W. FORREST 255
Numerical relations, similarity, and short-term recognition memory for pairs of digits. By W. A. WICKELGREN 263
The development of a keyboard training situation in miniature. By J. A. LEONARD 275
The effect of learning instruction on motor and verbal responses in recall. By F. J. TAYLOR 291
The responses of normal adult subjects to the matrices test. By N. E. WETHERICK 297
Perceptual sensitization, response suppression, and psychopathology. By T. M. CAINE 301
The size-constancy theory of visual illusions. By W. H. N. HOTOPF . . 307
Size constancy and intelligence: a re-examination. By V. HAMILTON . . 319
The effect of informational feedback on size judgements of schizophrenic patients. By R. STANNARD, G. SINGER and R. OVER 329
An experimentally induced conflict between vision and proprioception. By R. OVER 335
Effects of sequential ordering of added pairs and of correction *v* non-correction procedures on paired-associate performance. By W. F. BATTIG 343
Vanishing cues in paired-associate learning. By W. M. CHEYNE . . . 351
Transfer in category learning of young children: its relation to task complexity and overlearning. By A. M. CLARKE and G. M. COOPER 361
Avoidance responses as discriminated operants. By J. D. KEEHN . . . 375

PAGE

The effects of noise and doubling the signal frequency on individual differences in visual vigilance performance. By D. R. Davies and G. R. J. Hockey 381

Age differences in errors of commission. By G. S. Tune 391

Visual masking as a function of personality. By R. J. McLaughlin and H. J. Eysenck 393

A psycholinguistic investigation of a rule theory of 'sense'. By H. A. Alker. 397

Decoding a deceptive instruction. By S. Jones 405

Perceptual distortion induced by reasoning. By P. C. Wason and A. Kosviner 413

A reply to McPherson's comments on the Aberdeen conditioning studies. By H. J. Eysenck 419

Commentary on 'The gentic determination of differences in intelligence: a study of monozygotic twins reared together and apart' by Cyril Burt: congenital influences on the development of twins. By D. H. Stott . 423

Commentary on 'The genetic determination of differences in intelligence: a study of monozygotic twins reared together and apart' by Cyril Burt. By D. G. Lewis 431

A note on order of recall in short-term memory. By M. J. A. Howe . . 435

Publications Reviewed 437

Other Publications Received 503

INDEX OF AUTHORS

PAGE
Alker, H. A. . . . 397
Anthony, W. S. . . . 53
Battig, W. F. . . . 343
Bell, C. R. . . . 155
Boe, E. E. . . . 161
Bruner, J. S. . . . 71
Burt, C. . . . 137
Butcher, H. J. . . . 129
Caine, T. M. . . . 301
Cheyne, W. M. . . . 351
Clarke, A. D. B. . . . 113
Clarke, A. M. . . . 121, 361
Cooper, G. M. . . . 113, 121, 361
Dale, H. C. A. . . . 35
Das, J. P. . . . 87
Davies, D. R. . . . 381
Dawes, R. M. . . . 77
Dixon, N. F. . . . 239
Eysenck, H. J. . . . 99, 107, 419
Forrest, D. W. . . . 255
Hamilton, V. . . . 319
Hasan, P. . . . 129
Heim, A. W. . . . 171
Henney, A. S. . . . 121
Hockey, G. R. J. . . . 381
Holding, D. H. . . . 17
Hotopf, W. H. N. . . . 307
Howe, M. J. A. . . . 435
James, M. . . . 7
Jahoda, G. . . . 193
Jones, S. . . . 405
Keehn, J. D. . . . 375
PAGE
Keene, A. G. . . . 93
Kennedy, R. A. . . . 93
Kinsbourne, M. . . . 7
Kogan, W. S. . . . 161
Kosviner, A. . . . 413
Leonard, J. A. . . . 275
Levelt, W. J. M. . . . 225
Lewis, D. G. . . . 431
Locke, E. A. . . . 203
Mair, J. M. M. . . . 187
Matthews, W. A. . . . 25
Over, R. . . . 229, 335
Pastore, N. . . . 201
Pritchatt, D. . . . 17
Satinder, K. P.. . . . 1
Singer, G. . . . 329
Stannard, R. . . . 329
Stott, D. H. . . . 423
Tajfel, H.. . . . 71
Taylor, A. . . . 255
Taylor, F. J. . . . 291
Terwilliger, M. . . . 201
Thompson, W. . . . 99
Tune, G. S. . . . 391
Warrington, E. K. . . . 7
Wason, P. C. . . . 413
Watts, A. N. . . . 155
Watts, K. P. . . . 171
Wetherick, N. E. . . . 1, 297
Wickelgren, W. A. . . . 263
Willett, R. A. . . . 107
Woodhead, M. M. . . . 45

INDEX OF PUBLICATIONS REVIEWED

Asratian, E. A. *Compensatory Adaptations, Reflex Activity and the Brain* 205
Astrup, C. *Pavlovian Psychiatry. A New Synthesis* 476
Atkinson, R. C., Bower, G. H. & Crothers, E. J. *An Introduction to Mathematical Learning Theory* 468
Baker, C. *Visual Capabilities in the Space Environment* 445
Banks, C. & Broadhurst, P. L. *Στεφανος: Studies in Psychology* 497
Banton, M. *Roles: An Introduction to the Study of Social Relations* 223
Barbara, D. A. (ed.). *New Directions in Stuttering: Theory and Practice* 224
Barron, F. *et al.* *New Directions in Psychology. Vol. II* 438
Berges, J. & Lezine, I. *The Imitation of Gestures* 441
Berlyne, D. E. *Structure and Direction in Thinking* 437
Bennett, B. M. & Horst, C. *Supplement to Tables for Testing Significance in a 2 × 2 Contingency Table* 470
Bieri, J. *et al.* *Clinical and Social Judgement* 494
Brain, Lord. *Speech Disorders* 472
Brain Research, Vol. 1, no. 1 461
Carroll, J. *Language and Thought* 458
Cockcroft, J. *The Organisation of Research Establishments* 466
Cohen, J. *Behaviour in Uncertainty and Its Social Implications* 465
Cohen, J. *Human Robots and Myth in Science* 492
DuBois, P. H. *An Introduction to Psychological Statistics* 209
Dunlop, A. B. & McCabe, S. *Young Men in Detention Centres* 222
Eiferman, R. *Studies in Psychology* 468
Eysenck, H. J. & Rachman, S. *The Causes and Cures of Neurosis* 217
Eysenck, S. B. G. *Junior Eysenck Personality Inventory* 452
Faure, H. *Hallucinations et Réalité Perceptive* 448
Festinger, L. *Conflict, Decision and Dissonance* 440
Flower, R. M., Gofman, H. F. & Lawson, L. I. *Reading Disorders: a Multidisciplinary Symposium* 490
Forgus, R. H. *Perception* 499
Foulds, G. A. *Personality and Personal Illness* 451
Fox, M. V. *Canine Behaviour* 447
Freeman, T., Cameron, J. L. & McGhie, A. *Studies of Psychosis* 475
Gellhorne, E. & Loofbourrow, G. N. *Emotions and Emotional Disorders* 444
Gibson, H. B. *Gibson Spiral Maze Test* 471
Graham, C. H. *et al.* *Vision and Visual Perception* 485
Gregory, R. L. *Eye and Brain. The Psychology of Seeing* 462
Grygier, T. *et al.* *Criminology in Transition: Essays in honour of Mannheim* 449
Haber, R. N. *Current Research in Motivation* 487
Hambling, J. & Hopkins, P. *Psychosomatic Disorders in Adolescents and Young Adults* 219
Hammond, K. R. (ed.). *The Psychology of Egon Brunswick* 498
Hare, E. H. & Shaw, G. K. *Mental Health on a New Housing Estate* 474
Helson, H. *Adaptation Level Theory* 461
Herrnstein, R. J. & Boring, E. G. *A Source Book in the History of Psychology* 455
Hilgard, E. R. *Hypnotic Susceptibility* 477
Hochberg, J. E. *Perception: Foundations of Modern Psychology Series* 458
Holland, H. C. *The Spiral After-Effect* 464
Hollander, E. P. *Leaders, Groups and Influence* 441
Hopkins, P. & Wolff, H. *Principles of Treatments of Psychosomatic Disorders* 219
Horst, P. *Factor Analysis of Data Matrices* 454
Hotopf, W. H. N. *Language, Thought and Comprehension* 481
Howard, I. P. & Templeton, W. B. *Human Spatial Orientation* 500
Howells, J. G. *Modern Perspectives in Child Psychiatry* 215
Hull, C. L. *A Behavior System* 459
Hyman, R. *The Nature of Psychological Inquiry* 458
Isaacson, R. L., Hutt, M. L. & Blum, M. L. *Psychology: The Science of Behavior* 496
Johnson, R. C. & Midinnus, G. *Child Psychology. Behaviour and Development* 489
Kellmer Pringle, M. L. *Deprivation and Education* 214
Kendler, H. *Basic Psychology* 456
Kraupl-Taylor, F. *Psychopathology: Its Causes and Symptoms* 493
Lane, H. & Bem, D. *A Laboratory Manual for the Control and Analysis of Behaviour* 497
Lindren, H. C., Byrne, D. & Petrinovich, L. *Psychology: An Introduction to a Behavioural Science* 501
Ljung, B.-O. *The Adolescent Spurt in Mental Growth* 212
Luce, R. D., Bush, R. R. & Galanter, E. (eds.). *Handbook of Mathematical Psychology* 208
Lynn, R. *Attention, Arousal and the Orientation Reaction* 460
McGill, T. E. (ed.). *Readings in Animal Behaviour* 487
Macfarlane Smith, I. *Spatial Ability* 443
McNeil, E. R. *et al.* *The Nature of Human Conflict* 439

Maier, H. W. *Three Theories of Child Development* 472
Maier, N. R. F. *Psychology in Industry* 478
Malpass, L. F., Hocutt, M. O., Martin, E. P. & Givens, P. R. *Human Behavior: A Program for Self-Instruction* 479
Marks, I. M. *Pattern of Meaning in Psychiatric Patients* 476
Massarik, F. & Ratoosh, P. *Mathematical Explorations in Behavioural Science* 486
Meister, D. & Rabideau, G. *Human Factors Evaluation in System Development* 469
Merleau-Ponty, M. *The Structure of Behaviour* 480
Munn, N. L. *The Evolution and Growth of Human Behavior* 489
Munn, N. L. *Psychology: The Fundamentals of Human Adjustment* 501
Myers, J. K. & Roberts, B. H. *Family and Class Dynamics in Mental Illness* 221
Osgood, C. E. & Sebeck, T. A. *Psycholinguistics: A Study of Theory and Research Problems* 211
Paul, G. L. *Insight vs. Desensitization in Psychotherapy* 495
Porteus, S. D. *Porteus Maze Tests: 50 Years Application* 470
Problems of Television Research. Television Research Committee 492
Quay, H. C. *Juvenile Delinquency: Research and Theory* 449
Ratliff, F. *Mach Bands* 446
Reed, E. W. & Reed, S. C. *Mental Retardation: A Family Study* 453
Rimland, B. *Infantile Autism: The Syndrome of its Implications for a Neural Theory of Behaviour* 216
Rosen, E. R. & Gregory, I. *Abnormal Psychology* 219
Russel, R. (ed.). *Frontiers in Psychology* 439
Sandström, C. L. *The Psychology of Childhood and Adolescence* 473
Sarason, I. G. *Personality: An Objective Approach* 502
Schein, E. H. *Organisational Psychology* 450
Schultz, D. P. *Sensory Restriction: Effects of Behaviour* 463
Scott, J. P. & Fuller, J. L. *Genetics and the Social Behaviour of the Dog* 447
Shaw, O. L. *Maladjusted Boys* 214
Sherif, C. W., Sherif, M. & Nebergall, R. E. *Attitude and Attitude Change* 442
Shor, R. E. & Orne, M. T. (eds.). *The Nature of Hypnosis* 477
Smith, B. *Memory* 482
Spigel, I. M. (ed.). *Readings in the Study of Visually Perceived Movement* 464
Steinberg, H. (ed.). *Animal Behaviour and Drug Action* 206
Stephens, J. M. *The Psychology of Classroom Learning* 439
Stilson, D. W. *Probability and Statistics in Psychological Research and Theory* 486
Strange, J. R. *Abnormal Psychology: Understanding Behaviour Disorders* 218
Swets, J. S. *Signal Detection and Recognition by Human Observers* 483
Talland, G. A. *Deranged Memory: A Psychonomic Study of the Amnesic Syndrome* 207
Taylor, C. W. T. *Widening Horizons in Creativity* 467
Taylor, I. G. *Neurological Mechanisms of Hearing and Speech in Children* 471
Toch, H. *The Social Psychology of Social Movements* 491
Tomkins, S. S. & Izard, C. E. (eds.). *Affect, Cognition and Personality: Empirical Studies* 444
Vernon, J. *Inside the Black Room: Studies in Sensory Deprivation* 210
Watson, R. I. *Psychology of the Child* 212
Welford, A. T. & Birren, J. E. *Behaviour, Aging and the Nervous System* 491
Whittaker, J. O. *Introduction to Psychology. Students Manual to Psychology* 457
Willett, T. C. *Criminal on the Road* 222
Williams, M. *Mental Testing in Clinical Practice* 220
Wisdom, J. O. & Wolff, H. *The Role of Psychosomatic Disorder in Adult Life* 219
Wolman, B. B. *Handbook of Clinical Psychology* 474
Wolman, B. B. & Nagel, E. *Scientific Psychology* 210
Woodson, W. E. & Conover, D. W. *Human Engineering Guide for Equipment Designers* 479
Young, M. *Innovation and Research in Education* 213
Zubin, J. *et al. An Experimental Approach to Projective Techniques* 493

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 1–5
Printed in Great Britain

EFFECTS OF INTERMODAL STIMULATION ON FIGURAL AFTER-EFFECTS

BY K. PAUL SATINDER
*Department of Psychology, Panjab University, Chandigarh, India**

Verification was sought for the assumption by Kohler & Wallach that figural after-effect is a phenomenon specific to any given sense modality. Results contradicted Kohler & Wallach and Jaffe, as concurrent stimulation in different sense modalities affected significantly the figural after-effects occurring in other sense modalities in the absence of any size relation between the stimuli. The results support the hypothesis of sensory interaction, suggested by Soviet researches.

Jaffe (1956) questioned the assumption of the isomorphic representation of electrotonic activity in the primary projection area as put forth by Köhler & Wallach (1944), on the basis that 'if satiation is restricted to the primary projection areas, figural after-effects should only be demonstrable when both the antecedent stimulus and the stimuli used to test for the after-effect are presented to the same sense modality' (p. 71). He found that the 'concurrent visual stimulation can induce a significant kinaesthetic figural after-effect, but this after-effect appears only when contrasting size relationships exist between the visual stimuli' (p. 73). Research on sensory interaction conducted in the Soviet Union reviewed by London (1954) has clearly demonstrated modification of the response in one sense modality under the influence of stimulation of another sense modality, without any size relations between the stimuli. The present investigation was conducted to determine the effects of intermodal stimulation upon figural after-effects in the auditory, kinaesthetic and visual sense modalities.

METHOD

Design of the experiment

The investigation required the following conditions for each sense modality. (i) Basal or pre-inspection point of subjective equality (PSE), to serve as the control measure (McEwen & Rodger, 1960). (ii) Post-inspection PSE, after an inspection period of 1 min (Christman, 1954; Bourne & Beier, 1961; Hammer 1949). (iii) Post-inspection PSE, after an inspection period of 1 min, with concurrent stimulation in one or other of the remaining sense modalities. (iv) Post-inspection PSE, after an inspection period of 1 min with concurrent stimulation in both the other sense modalities.

Concurrent stimulation was given during inspection period only. The T-figures of the respective sense modalities were used for concurrent stimulation.

The fulfilment of these conditions required the following experimental schedule (cf Table 1).

All the PSEs were based upon four judgements (Christman, 1954). Each judgement was followed by a 2 min rest period, determined on the basis of a pilot study. The method of limits was used with counterbalancing (Christman, 1954; Satinder, 1963). All the conditions were presented to each subject in different random orders, as it tended to reduce individual differences. On the basis of a pilot study (Satinder, 1963), a rest period of 5 min was allowed between conditions.

---

* Now at Department of Psychology, University of Mysore, Manasa Gangotri, Mysore-2, India.

Apparatus

(*a*) *Auditory after-effects.* An audio-generator, amplifier, matched headphones and a system of relays and switches provided manual control of stimulus duration, phone selection and pitch at the two ears.

(*b*) *Kinaesthetic after-effects.* The inspection (I) and test (T) figures were 2·5 and 1·5 in. in width respectively. These were made of aluminium but were otherwise the same as those used by Jaffe (1956), and by McEwen & Rodger (1960). An adjustable comparison scale (Satinder, 1964) was used for measuring kinaesthetic figural after-effects.

(*c*) *Visual after-effects.* The stimulus pattern of Fig. 1 was used. The inspection figure was a ring of light, 40 mm in inner and 43 mm in outer diameter. The fixation mark, a small red dot of light, was 10 cm to the right of the centre of the I-figure as viewed by the subject. The test figures were two disks of light varied in size by means of iris diaphragms. The figures were lit by a lamp concealed behind the screen. The fixation point was between the two disks, 10 cm from the centre of each. The standard T-figure (falling inside the area of the I-figure) was on the subject's left and remained constant at 20 mm diameter. The neutral T-figure was varied by the experimenter and its size could be measured from the scale at the back of the screen on which the diaphragm was mounted.

Table 1. *Experimental schedule of PSE determinations*

Condition for which PSE determined	Abbreviations
Basal—for auditory perception	—
Basal—for kinaesthetic perception	—
Basal—for visual perception	—
After auditory inspection, i.e. auditory figural after-effect	AFAE
After kinaesthetic inspection, ie. kinaesthetic figural after-effect	KFAE
After visual inspection, i.e. visual figural after-effect	VFAE
After auditory stimulation with visual inspection	A-VFAE
After kinaesthetic stimulation with auditory inspection	K-AFAE
After visual stimulation with kinaesthetic inspection	V-KFAE
After auditory stimulation with kinaesthetic inspection	A-KFAE
After kinaesthetic stimulation with visual inspection	K-VFAE
After visual stimulation with auditory inspection	V-AFAE
After auditory and kinaesthetic stimulation with visual inspection	AK-VFAE
After kinaesthetic and visual stimulation with auditory inspection	KV-AFAE
After visual and auditory stimulation with kinaesthetic inspection	VA-KFAE

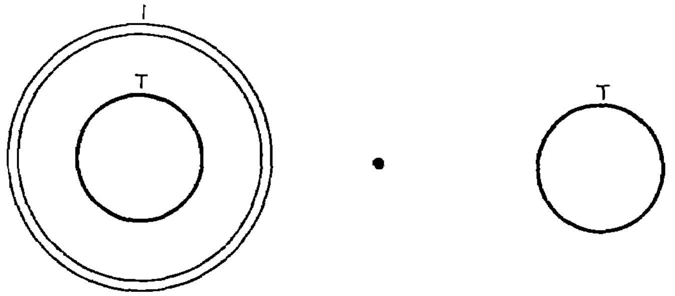


Fig. 1. Stimulus pattern used for visual figural after-effects (cf. Köhler, 1951, p. 203).

Subjects

Thirty postgraduate students, fifteen of each sex, served as subjects (mean age 22·4 yr; range 18–32 yr). None of the subjects was familiar with figural after-effects.

Procedure

The experiment was conducted in a dark room. The following procedure was adopted in testing the subjects individually.

(*a*) *Auditory figural after-effects.* The general procedure of Christman (1954, p. 485) was followed. Each subject was given the following instructions: 'A tone will be presented to the left ear for some time. Immediately after the termination of that tone another will be presented to the right ear and will be varied and you are required to inform me when you feel that it is just equivalent to the preceding one to the left ear.'

The basal PSE was first determined. The standard 600 c/s tone, i.e. the test figure, was presented to the left ear for 2 sec. After a 2 sec pause, a measuring tone of 500 or 700 c/s was presented to the right ear, with slight variations from trial to trial. The experimenter proceeded to adjust this tone to the point of subjective equality, as reported by the subject.

An 800 c/s tone was then presented to the left ear for 1 min as 'inspection figure', and a post-inspection PSE was determined in the same manner as the basal PSE.

(*b*) *Kinaesthetic figural after-effects.* The subject was seated on an adjustable stool, between two tables, 34 in. in height (Bakan, Myers & Schoonard, 1962) on which the figures were mounted. Inspection and test figure were placed 3 in. apart on the right-hand side, the T-figure being nearer to the subject and the comparison figure 22 in. away from the T-figure on the left-hand side of the subject (Köhler & Dinnerstein, 1947; Bourne & Beier, 1961; Heinemann, 1961). Fig. 2 illustrates the arrangement of the apparatus for kinaesthetic figural after-effects. All the three figures were parallel to one another (Bakan *et al.* 1962) and at the same level from the ground (Nachmias, 1953). The post-inspection PSE was determined after the subject had inspected the I-figure by making sixty movements in 1 min at a rate of 1/sec with the help of an electric metronome (Satinder, 1963).

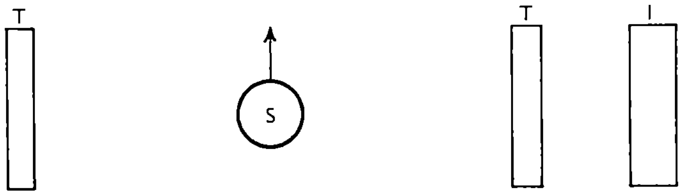


Fig. 2. Arrangement of the apparatus for kinaesthetic figural after-effects. The arrow indicates the direction of the subject's line of sight.

(*c*) *Visual figural after-effects.* The subject was seated on the adjustable stool, at a distance of 100 cm from the apparatus. The T- and I-figures were at the subject's eye level. For determining the pre- and post-inspection PSEs, the procedure of McEwen & Rodger (1960) was followed except that the method of limits was used, not the constant method.

RESULTS

The PSEs for all conditions and all sense modalities were determined for each subject. Hence there were fifteen PSEs for each subject: five under each sense modality (cf. Table 1). To make the data comparable, the PSEs for the after-effects were converted into percentage change from the basal PSE value of the corresponding sense modality (all of the PSE values for after-effects have been expressed as percentage *decreases* from the corresponding basal value since they were all less than the corresponding basal). Averages were computed for each of the conditions based on the scores of the thirty subjects: *t*-tests were used to test differences (cf. Tables 2, 3 and 4).

The auditory figural after-effect is significant ($P < 0{\cdot}001$). Concurrent kinaesthetic and visual stimulation, singly as well as combined, significantly increased the after-effect, but visual stimulation did so more than kinaesthetic stimulation. Their combined effect is again larger than either alone; it differs significantly from kinaesthetic stimulation alone ($P < 0{\cdot}01$), but not from visual stimulation alone.

The kinaesthetic figural after-effect is significant ($P < 0{\cdot}001$). Concurrent visual and auditory stimulation, singly and in combination, increased the after-effect; visual stimulation did so significantly ($P < 0{\cdot}02$) but not auditory stimulation. The combined effect is again larger than the single ones and differs significantly from auditory stimulation alone ($P < 0{\cdot}05$) but not from visual stimulation alone.

Table 2. *Means for auditory figural after-effects and their modification by stimulation in other sense modalities, and* t*-ratios of the differences between means*

Condition		Basal	AFAE	K-AFAE	V-AFAE	KV-AFAE
	Means	0·00	4·97	6·08	6·66	7·76
Basal	0·00	—	12·743*	—	—	—
AFAE	4·97	—	—	2·181****	2·091****	3·902*
K-AFAE	6·08	—	—	—	—	2·809**
V-AFAE	6·66	—	—	—	—	1·641
KV-AFAE	7·76	—	—	—	—	—

* $P < 0{\cdot}001$. ** $P < 0{\cdot}01$. **** $P < 0{\cdot}05$.

Table 3. *Means for the kinaesthetic figural after-effects and their modification by stimulation in other sense modalities, and* t*-ratios of the differences between means*

Condition		Basal	KFAE	V-KFAE	A-KFAE	VA-KFAE
	Means	0·00	9·55	11·86	11·56	13·19
Basal	0·00	—	12·733*	—	—	—
KFAE	9·55	—	—	2·481***	1·974	3·497**
V-KFAE	11·86	—	—	—	—	1·474
A-KFAE	11·56	—	—	—	—	2·327****
VA-KFAE	13·18	—	—	—	—	—

* $P < 0{\cdot}001$ ** $P < 0{\cdot}01$ *** $P < 0{\cdot}02$ **** $P < 0{\cdot}05$

Table 4. *Means for the visual figural after-effects and their modification by stimulation in other sense modalities, and* t*-ratios of the difference between means*

Condition		Basal	VFAE	A-VFAE	K-VFAE	AK-VFAE
	Means	0·00	4·91	5·90	5·85	7·19
Basal	0·00	—	12·589*	—	—	—
VFAE	4·91	—	—	3·113**	3·197**	4·662*
A-VFAE	5·90	—	—	—	—	3·341**
K-VFAE	5·85	—	—	—	—	3·221**
AK-VFAE	7·19	—	—	—	—	—

* $P < 0{\cdot}001$ ** $P < 0{\cdot}01$

The visual figural after-effect is significant ($P < 0{\cdot}001$). Concurrent auditory and kinaesthetic stimulatin, both singly and combined, significantly increased the after-effect, auditory stimulation did so more than the kinaesthetic stimulation; the combined effect is again significantly larger than the single ones.

In general kinaesthetic figural after-effects were more marked than auditory and visual figural after-effects.

Discussion

The figural after-effects in the auditory, kinaesthetic and visual sense modalities are in the predicted direction (Christman, 1954; Köhler & Dinnerstein, 1947; Köhler & Wallach, 1944; Köhler, 1951). Further, the results tend to indicate that the figural after-effect in one sense modality is significantly affected by concurrent stimulation in other sense modalities. This finding, the possibility of which was envisaged by Köhler (1951), is contrary to the view that figural after-effects are specific to particular sense modalities. The results contradict Jaffe's (1956) conclusion, that concurrent stimulation would influence after-effects only when there were contrasting size-relationships between the stimuli. But they are in agreement with the results of research on sensory interaction in Soviet Union, reviewed by London (1954).

REFERENCES

BAKAN, P., MYERS, L. B. & SCHOONARD, J. (1962). Kinesthetic after-effects and length of inspection-period. *Amer. J. Psychol.* **75**, 457–61.

BOURNE, L. E. JR. & BEIER, E. G. (1961). Effects of duration of inspection upon kinesthetic figural after-effects. *J. gen. Psychol.* **65**, 163–9.

CHRISTMAN, R. J. (1954). Shifts in pitch as a function of prolonged stimulation with pure tones. *Amer. J. Psychol.* **67**, 484–91.

HAMMER, E. R. (1949). Temporal factors in figural after-effects. *Amer. J. Psychol.* **62**, 337–54.

HEINEMANN, E. G. (1961). Figural after-effects in kinesthetic effects of object width and repeated presentations. *J. exp. Psychol.* **61**, 51–6.

JAFFE, R. (1956). The influence of visual stimulation on kinesthetic figural after-effects. *Amer. J. Psychol.* **69**, 70–5.

KOHLER, W. (1951). Relational determination in perception. In *Cerebral Mechanisms in Behavior: the Hixon Symposium* (ed. L. A. Jeffress). New York: Wiley.

KÖHLER, W. & DINNERSTEIN, D. (1947). Figural after-effects in kinesthesis. In *Miscellanea psychologica*, ed. A. Michotte. pp. 196–220. Louvain: Publications Universitaires.

KÖHLER, W. & WALLACH, H. (1944). Figural after-effects: an investigation of visual processes. *Proc. Amer. phil. Soc.* **88**, 269–357.

LONDON, I. D. (1954). Research on sensory interaction in the Soviet Union. *Psychol. Bull.* **51**, 531–68.

MCEWEN, P. & RODGER, R. S. (1960). Some individual differences in figural after-effects. *Brit. J. Psychol.* **51**, 1–8.

NACHMIAS, J. (1953). Figural after-effects in kinesthetic space. *Amer. J. Psychol.* **66**, 609–12.

SATINDER, K. P. (1963). Intermodality interaction of figural after-effects and their relationship with personality and sex. Unpublished Master's Thesis, Panjab University, India.

SATINDER, K. P. (1964). Another device for the study of kinesthetic figural after-effects. *J. psychol. Res.* **8**, 92–3.

(*Manuscript received* 23 *March* 1964; *revised manuscript received* 30 *August* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 7–16
Printed in Great Britain

UNCERTAINTY AND TRANSITIONAL PROBABILITY IN THE SPAN OF APPREHENSION

BY ELIZABETH K. WARRINGTON, M. KINSBOURNE AND MERLE JAMES

The National Hospital, Queen Square, London and Institute of Experimental Psychology, Oxford

Span of apprehension was measured for stimulus ensembles of various sizes, using letters, digits and an arbitrarily selected set of simple lines. Span was significantly decreased when letters and digits were presented as mixed sequences, in proportion to the number of letter-digit juxtapositions. Span also varied with the serial position at which a single juxtaposition occurred, being greatest where this was near the mid-point. These results were thought to arise from the higher transitional probabilities between items of the same class than of different classes, and the greater ease in grouping such items. No effect of stimulus uncertainty was found. Random letter span was significantly less than digit span, but span for sequences reflecting letter frequencies in the language did not differ significantly from digit span (at 100 msec exposure duration).

The unfamiliar line material yielded relatively lower spans, varying inversely with ensemble size (two, four and eight choice). This was determined not by overall stimulus uncertainty but by the degree of heterogeneity of the individual stimulus arrays. The relatively less heterogeneous stimulus assays were thought to lend themselves more readily to verbal recoding. With very familiar material transitional probability, in relation to previous experience of the language, determines span. With the less familiar material not subject to such response bias, it was the recodability of the individual display which determined the span. With only a single exception, comparable results were obtained at moderately and at very brief exposure durations.

Span of apprehension and memory span have repeatedly been found to be greater for digits than for letters (Cattell, 1886; Jacobs, 1887; Crossman, 1961; Mackworth, 1963). This could be an effect of ensemble size (ten as against twenty-six alternatives) as demonstrated for reaction time (Hick, 1952; Hyman, 1953) and for span of apprehension with other material (Krulee, Podell & Ronco, 1954; Fraisse & Blancheteau, 1962). However, the naming of digits and letters is highly familiar and in such tasks the effects of stimulus uncertainty may not be demonstrable (Crossman, 1953; Schmidtke, 1961; Davis, Moray & Treisman, 1961). Other explanations for the digit and letter span disparity might take account of differences in sequential dependency between successive items, in the number of individual items repeated in a single display, and in the greater probability of a random guess being correct when the ensemble is small. The present investigation attempts to determine the relative contribution made by factors such as these to the disparity between digit and letter span of apprehension, and applied similar tests to perceptually simple but relatively unfamiliar stimulus material consisting of straight and curved lines in various orientations.

Sperling (1960) claimed that span of apprehension is limited by the capacity of the immediate memory store. Even if this is so at conventional exposure durations, it could hardly be true at exposure durations just above threshold, when letter span is substantially diminished (Mackworth, 1962) and the strain on immediate memory must be less. For this reason all measures were here carried out both at the conventional exposure duration of 100 msec and at very brief exposures just above threshold duration.

The classical guessing correction (Guilford, 1936) assumes that incorrect responses are random. This assumption is not entirely justified for recognition of briefly presented alphabetical material (Bricker & Chapanis, 1953; Conrad & Hull, 1964). However no alternative correction can be adapted to the data here obtained, and the classical correction has been applied to all results on the view that more reliance can be placed on differences in span for different materials if they remain significant after the application of a guessing correction.

There were three experiments. Expt. I examined the effect on span of apprehension of varying ensemble size and the expectancy preceding each test exposure. Expt. II developed a finding from the first experiment in relation to span for alphanumerical strings. Expt. III applied methods comparable to those of the first experiment to a less familiar stimulus set.

METHOD

The following details of method applied to all three experiments.

Apparatus

An electronic tachistoscope with test field subtending 14° × 9° of visual angle was viewed binocularly by subjects in the light-adapted state. The illumination of the test field and pre- and post-exposure fields was held constant at 2·5 ft.-Lamberts. Exposure duration of the test stimulus could be varied between 2 and 1600 msec, intermediate steps providing increments of 25 %. Further details of the apparatus have been given by Kinsbourne & Warrington (1962).

Test materials

Capital letters and digits were printed and linear forms were drawn in black on white cards. The letters were capitals, omitting 'I' and 'O'. The digits were 0–9. Twenty-four cards each had a single capital letter, and eight each had one of the lines, in the centre. These were used for threshold measurements. Letter and digit strings always consisted of six items, horizontally closely adjacent. Each item subtended $\frac{1}{2}$° at the eye. The strings were used for measuring the span of apprehension. Each item was drawn out of a pool of all possible alternatives, in such a way that it appeared with equal frequency in each position in any set of strings. Chance determined whether an individual item would recur within the same string. Strings so prepared are referred to as random. Only in Expt. II were sequences prepared by any other method.

Subjects

The subjects were unpaid volunteers in the hospital service, aged 18–50, inexperienced in span of apprehension experiments, and unaware of the purpose of the work.

Procedure

The testing began with a threshold measurement for single letters. Five single centrally placed letters were consecutively exposed at 2 msec, and again five, randomly selected, at durations representing successive increments of 25 % until a set of five letters was correctly identified. An exposure duration 25 % longer than that so established was then defined for present purposes as 'threshold' for graphic symbols. A comparable method was used to establish 'threshold' for the line material.

The relevant experiment was then carried out in a single session. Exposures were always in two sets, at 100 msec and at 'threshold' duration respectively. The subject was asked to fixate a faint centrally placed cross in the pre-exposure field and warned to be 'ready' before each exposure. He was instructed to report the items he saw, in the order in which they appeared. Responses were spoken in the case of the letters and digits and were recorded by the experimenter. With the lines, responses were in writing, on pages ruled with five columns, representing the five positions of the items in the sequence.

The separate sets in each experiment were presented in counterbalanced order within the group of subjects to obviate differential practice effects.

The number of items correctly reported in the right position within the sequence was counted. The result was subjected to a classical guessing correction (Guilford, 1936), taking into account number of alternatives and number of 'false positives'. The figures, uncorrected and corrected, were transformed into percentages.

EXPERIMENT I

The purposes of the experiment were to determine span of apprehension for numerical and alphabetical material and to relate it to (i) ensemble size, (ii) conformity and nonconformity of the frequency distribution of letters to that in the language (iii) pre-stimulus cueing as to class of stimulus array, (iv) test exposure duration.

To examine the effect of ensemble size on span of apprehension, with letters and with letters and digits as test items, two further sets were constructed. One was small, the five vowels, one large, the letters and digits combined (34 in all, omitting I and O). Span for these materials was compared with standard letter and digit span. A further variable was expectancy preceding each individual trial. In one type of presentation the subject was told what class of material to expect at each exposure, in the other he was not.

Test materials

The following materials were used: (*a*) random digit strings (20 test cards); (*b*) random letter strings (24 test cards); (*c*) random 'alphanumerical' (mixed) strings (34 test cards); (*d*) random vowel strings (20 test cards); (*e*) zero-order letter strings (15 test cards); (*f*) first-order letter strings (15 test cards); (*e*) and (*f*) were derived from Miller, Bruner & Postman (1954) using the first six items of each of their eight-letter sequences.

Experimental plan

Three groups of subjects were used and thirty-six subjects took part.

In group I twelve subjects were shown *a*, *b* and *c* separately (with pre-stimulus cueing as to class) in counterbalanced sequence, and six subjects were shown *a*, *b* and *c* combined in random sequence, without prestimulus cueing.

In group II six subjects were shown *a*, *b* and *d* separately, in counterbalanced sequence.

In group III twelve subjects shown *a*, *e* and *f* separately, in counterbalanced sequence.

In each section, half the subjects were tested first at 'threshold', the others were tested first at 100 msec.

Results

The percentage mean spans, before and after guessing correction, at each of the two exposure durations, for letter, digit, alphanumerical (mixed), vowel, zero-order and first-order strings, are given in Table 1 and the significance of the differences in Table 2. Digit span significantly exceeds letter span, which in turn significantly exceeds mixed span.

Vowel span did not differ significantly from letter span. Digit span significantly exceeded zero order letter span It did not differ significantly from first-order letter span at 100 msec, but just exceeded it at threshold ($P < 0{\cdot}05$). Foreknowledge of the class of test material to be expected at each presentation did not diminish the significance of these differences between the letter and the digit spans. For each of the six sets of test material, separately presented, span was significantly higher for the higher exposure duration

Table 1. *Mean percentage scores for span of apprehension for letters, digits, vowel and mixed arrays in Expt.* 1.

Group I, $n = 18$	T		100	
	S, $n = 12$	C, $n = 6$	S, $n = 12$	C, $n = 6$
(*a*) Digits	**67** 68	**68** 68	**76** 77	**79** 79
(*b*) Letters	**58** 58	**55** 55	**67** 67	**68** 68
(*c*) Mixed	**52** 52	**53** 53	**61** 61	**68** 68

Group II, $n = 6$	T	100
	S	S
(*a*) Digits	**56** 57	**73** 74
(*b*) Letters	**49** 50	**64** 64
(*d*) Vowels	**52** 56	**66** 69

Group III, $n = 12$	T	100
	S	S
(*a*) Digits	**74** 75	**87** 87
(*e*) Zero order	**64** 64	**78** 79
(*f*) First order	**69** 69	**85** 85

The abbreviations used in the table are 'T' for threshold exposure, '100' for 100 msec exposure, 'S' for separate sequences and 'C' for combined sequences. Percentage scores corrected for guessing are in bold type.

Table 2. *Comparisons between mean percentage scores for the conditions of Expt. I* (t *values*)

	Letters				Digits				
	Threshold		100 msec		Threshold		100 msec		
	S	C	S	C	S	C	S	C	Exposure *t* against 100
(*a*) Digits	7·17***	4·07**	5·07***	5·11**	—	—	—	—	5·99***
(*b*) Letters	—	—	—	—	7·17***	4·07**	5·07**	5·11**	6·40***
(c) Mixed	5·14***	1·02	4·24**	0·75	10·60***	5·92***	14·20***	5·65***	5·30***
(*d*) Vowels	0·47	—	0·82	—	1·58	—	4·54**	—	7·04***
(*e*) Zero order	—	—	—	—	4·19**	—	5·38***	—	5·58***
(*f*) First order	3·68*	—	3·19*	—	2·72*	—	0·60	—	4·87***

*** $P < 0{\cdot}001$ ** $P < 0{\cdot}01$ * $P < 0{\cdot}05$

($P < 0{\cdot}001$). The comparisons between performances on the various test materials (Table 2) yielded the same results whether the presentations were at 'threshold' or at 100 msec, with the one exception mentioned above.

EXPERIMENT II

In the previous experiment span for mixed letters and digits was found to be significantly lower than for either digit or letter sequences. This could be due either to the greater stimulus uncertainty involved or to the decreased transitional probability at letter–digit justapositions. The second of these possibilities was examined in this experiment. The purposes of the experiments were to compare the span of apprehension for mixed strings containing letters and digits (i) in relation to the number of letter-digit juxtapositions and (ii) in relation to the serial position of a single juxtaposition.

Test materials

Two sets of test materials were prepared: In set A the proportion of letters and digits within any one string was held constant at 3:3; the relative positions of the items of the two classes were arranged so that the number of letter–digit juxtapositions varied between 1 and 5 (cf. Table 3). In set B, twenty strings of mixed letters and digits were so arranged that only at one (variable) point would letter and digit be adjacent; the transition could come after the first, third, fourth or fifth symbol. There were four mixed strings for each position of transition between classes; in two of them digits preceded letters and in two letters preceded digits.

Subjects

Sixteen subjects were tested with set A and twelve subjects with set B.

Results

Effect of number of transitions between classes. Scores at the two exposure durations were summed for each subject. Table 3 shows the mean percentage score for each condition. With increasing number of switches between classes the span of apprehension progressively decreased. An analysis of variance showed that the decrease in total span with the increase in number of switches is significant ($F = 9{\cdot}77$; D.F. 4, 15; $P < 0{\cdot}01$). Mean percentage scores differing by more than 3·3 % differ significantly at $P < 0{\cdot}05$.

Effect of position of transition between classes. The span for strings of each position of transition between classes is given in Table 3 (scores at the two exposure durations were summed). Analysis of variance showed that the position of the transition affects the span of apprehension significantly ($F = 4{\cdot}95$; D.F. 4, 11; ($P < 0{\cdot}05$). Mean percentage scores differing by more than 4·6 % differ significantly at $P < 0{\cdot}05$. There is a significantly smaller span when the break is in positions 1, 2 and 5 than when it is in position 3 in the centre of the string. Position 4 gives an intermediate score.

The results show that span drops as the number of letter-digit alternations increase, and the transitional probabilities within the sequences consequently decrease. When the number of alternations is held constant, at one, the position of alternation has a significant effect, yielding a greater span when it is centrally rather than eccentrically placed. This may be because when items are grouped (into 'chunks') for storage in short-term memory, such groups are more readily formed from items of homogeneous class.

Table 3. *Mean percentage score related to number and position of change in class (Expt. II)*

No. of juxtapositions	No. of cards	Mean score	Position of breaks	Mean score
1	4	50·4	1	62·8
2	8	47·6	2	62·8
3	12	44·2	3	69·4
4	4	43·9	4	66·7
5	4	41·9	5	62·7

EXPERIMENT III

The purpose of the third experiment was to measure span of apprehension in relation to ensemble size for simple but unfamiliar material.

Test materials

Simple straight and curved lines in various orientations were used as test material in this experiment. Three sets of sixteen cards were prepared, each bearing five items. The sets included items drawn from two, four and eight alternatives respectively. The two-choice set was in turn divided into two subsets of eight cards. Each subset comprised items drawn from different pairs of alternatives. The ensembles were as follows:

Two choice: (I*a*) ∩)
(I*b*) / (
Four choice: (II) — / (∪
Eight choice: (III) — / \ ∩ ∪ ( )

Subjects

Twenty-six subjects took part in the experiment. All were shown the stimulus cards at 100 msec. Eight of them were also shown the material at 'threshold'.

Results

There was a significant decline in span as the number of alternatives was increased, from two to four and then to eight, both before and after making guessing corrections (cf. Table 4). When subsets I*a* and I*b* were compared, the span on the latter was significantly the greater.

Table 4. *Comparison of scores for line sequences in relation to size of set of alternatives* (*Expt. III*)

Set	Mean scores					*t* values	
No. of alternatives	'Threshold'		100 msec		Comparison	'Threshold'	100 msec
I*a* 2	**71·8**	87·6	**80·4**	81·2	I*a* and I*b*	3·11*	3·31**
I*b* 2	**85·0**	90·0	**92·9**	95·2	I*a* and II	2·53*	3·62**
I 2	**78·4**	85·6	**86.6**	91·1	I and II	5·68***	4·33***
II 4	**60·0**	70·2	**71·2**	78·6	II and III	0·71	3·20**
III 8	**58.0**	62·6	**62.7**	66·4			

Scores corrected for guessing are given in bold type.
$^{***}P < 0{\cdot}001$ $^{**}P < 0{\cdot}01$ $^{*}P < 0{\cdot}05$

Table 5. *Scores on line sequences related to size of set of alternatives and number of different items per sequence* (*Expt. III*)

Different items per card	Set II	Set III
2	83·7	81·1*
3	78·8	
4	74·8	69·6
5	—	56·7

* There were only three cards in these categories, insufficient for statistical analysis.

In each set, chance factors determined the number of different figures actually to appear in any individual stimulus array. While two figures occurred on all cards of set I, those of set II could be classified into cards bearing two, three and four different figures, while set III also included cards in which figures in all five serial positions were different. This made it possible to compare performance on strings of comparable complexity under different conditions of expectancy, and conversely, of unequally complex material presented under comparable conditions of expectancy (cf. Table 5).

An analysis of the results led to the following findings. In set II there was a significantly lower performance on the subset with three different figures per card, compared with the subset with two ($t = 2{\cdot}64$, D.F. 25, $P < 0{\cdot}05$), though there was no significant difference between subsets with three and with four ($t = 1{\cdot}49$, D.F. 25, $P > 0{\cdot}05$). In set III, there was a significantly lower performance on the subset with five figures per card, as compared to that with four ($t = 5{\cdot}15$, D.F. 25, $P < 0{\cdot}001$).

When scores for cards bearing four different figures derived from set II were compared with scores for such cards in set III, no significant difference was found ($t = 1{\cdot}23$, D.F. 25, $P > 0{\cdot}05$).

Discussion

When simple but unfamiliar material was used to measure span of apprehension, performance appeared to be related to the size of the set of alternatives when this was varied between two and eight items. An analysis was then performed to establish whether this was due to the effect of differences in expectancy, or to other factors. Ensembles of different size necessarily give rise to sequences with overall differences in internal structure. These are based on the greater probability that an individual figure will recur within the same sequence when the number of alternatives is small. Such recurrence might simplify the response task by reducing the number of different responses that require activation. Further, there will be more sequences showing symmetry, and more in which a given figure is repeated in adjacent serial positions, which might make for ease in selection of the response. The analysis showed that differences in internal structure account for the difference in performance on sets II and III. The number of different figures within a sequence had an influence unconnected with the stimulus uncertainty. Set I was in two parts: in one instance (set I*a*) both figures were curved, though differently orientated; in the other (set I*b*) one figure was straight and one was curved, and again they differed in orientation. Scores were significantly higher for the second set, suggesting that the extra dimension of difference between the two figures, straight as against curved, was of help.

It appears, therefore, that where figures in an array differ in a more readily definable manner, span of apprehension is enhanced. This effect seems to depend on greater ease in recoding for short-term memory storage. Further, span is enhanced when, within a sequence of given length, the same figure appears more than once. Again, this may be due to greater ease in recoding such sequences.

DISCUSSION

The superiority of digit over letter span of apprehension (Cattell, 1886) was confirmed and shown to retain significance after application of a guessing correction. Reporting a comparable finding for memory span, Jacobs (1887) remarked on the relatively greater frequency, in everyday life, of apparently random digit sequences than of random letter sequences. He also pointed out that letters are derived from a larger ensemble than digits. Thus random letter sequences convey more information for two reasons: because of low transitional probability, and because of the larger size of the set of alternatives. The effects of transitional probability and of ensemble size were here tested in relation to letters and digits, as well as to non-symbolic stimulus material (with special reference to the discrepancy between digit and letter spans).

In the first experiment, stimulus uncertainty was varied by arbitrarily limiting ensemble size (to the five vowels) or extending it (letters and digits mixed), and by either providing or withholding prestimulus information as to the class of material to be expected. The latter procedure had no effect on span. The former failed to affect span in the case of vowels, but lowered span in the case of the mixed 'alphanumerical' strings. This result can also be explained on the view that transitional probabilities within homogeneous lists are greater than those within lists incorporating more than one class of material. In the second experiment this possibility was examined by varying the arrangement of items within mixed lists when their overall composition and therefore the stimulus uncertainty was held constant. In one test the number of transitions between classes was varied, and in the other the point within the sequence of a single such transition. A significant decrement of span was found as the number of digit-letter juxtapositions was increased. This could be related to the decrease in transitional probability brought about in this way. With a single transition only span was maximal where this divided the list into roughly equal parts. This may be because a break in this position makes for efficient 'chunking' in immediate memory (Miller, 1956), each chunk being homogeneous as to class, while an eccentric transition would give rise to lower transitional probabilities within the chunk that now has to contain a digit–letter transition.

Decreased memory span for mixed lists of letters and digits was found by Conrad (1962) and with dichotic presentation Yntema & Trask (1963) and Broadbent & Gregory (1964) found decrease in efficiency of report when the message to each ear contained items of more than one class. Basing themselves on evidence provided by Treisman (1960), Broadbent & Gregory (1964) suggest that transitional probabilities are reflected in a response set which has to be shifted, with consequent consumption of time, when classes are switched. This view can readily be applied to the present results.

The comparisons in the first experiment fail to reveal an effect of stimulus uncertainty, varied by changing size of set of alternatives, on span of apprehension. They also provide a way of accounting for the digit–letter span discrepancy in terms of transitional probability. First-order letter sequences yielded a span significantly in excess of zero order (in confirmation of Miller *et al.* 1954). Further, there was no significant difference between digit span and first-order letter span at 100 msec exposure duration: when the items in both the digit and the letter sequences were chosen with regard to their natural frequency (assuming the digits to be roughly equiprobable) the difference in span disappeared. This suggests that the discrepancy between letter and digit span of apprehension originally found was attributable to the difference in probability, in terms of previous experience, of randomly selected letter and digit strings.

Information may be varied by altering the number of equally probable alternatives from which the stimulus items are drawn, by varying the relative probability of these items, and by introducing sequential dependencies between them. Mathematical equivalence of these methods was demonstrated by Shannon & Weaver (1949) and this was shown to correspond to psychological equivalence in the case of reaction time by Hyman (1953). However, in the present series of experiments when familiar material and highly compatible responses were used, the artificial manipulation of these variables did not have an equally great effect. It appears that increasing redundancy of letter series effectively increases span only when this redundancy is that of the English language rather than an arbitrary redundancy determined by the experimental conditions.

Failure to demonstrate effects of stimulus uncertainty for highly compatible responses to familiar stimuli such as letters or digits cannot be used to discount expectancy effects for other materials, particularly where differences in transitional probability based on previous experience do not exist. As an instance of a simple but unfamiliar ensemble, a set of geometrical lines was used for span of apprehension measurements in the third experiment. These items were simple to recognize but less easily recorded verbally than are forms with symbolic connotation. Sets containing two, four and eight alternatives respectively were constructed. The stimulus cards used each bore five items, a number chosen on the basis of pilot studies which suggested that this would make for optimal performance on the four-choice and eight-choice ensembles. A significant inverse relationship was found between span and ensemble size. The responses were further examined to determine whether this result was indeed due to stimulus uncertainty or to factors related to the actual arrangement of items on the cards. It was found that within the eight-choice set there was a significant superiority of span on cards bearing four or fewer different items as compared to cards on which each of the five items was a different one. When this latter group was eliminated, scores on the four-choice and on the eight-choice sets no longer differed significantly. This shows that the difference originally found did not derive from the disparity in stimulus uncertainty for the two groups, but was related to the degree of heterogeneity of the array; that is, to the form rather than the amount of the redundancy (Garner, 1962).

The two-choice scores were derived from the use of two different stimulus pairs—one differing only in the orientation of two curved lines, the other differing also in that one line was straight, the other curved. Span was significantly greater for the set in which the two items differed more widely. This could be due to greater discriminability of the second pair, or to greater ease in recording the difference between the two items of this pair in verbal terms.

The span for the sets of four and of eight alternatives was shorter than the letter or the digit span, in spite of the simplicity of the stimuli, and the highly compatible response (manual imitation). Probably verbal recoding intervenes between stimulus and response for at least some of the simultaneously exposed items and this is more time-consuming for the non-symbolic line stimuli. Correspondingly, arrangements which might be expected to facilitate recoding, such as the recurrence of the same item in the series, should and do enhance performance.

Throughout this series of experiments, largely comparable results were obtained for exposures at 100 msec, and for exposures at just above the visual duration threshold. If different factors limit span at moderately and at exceedingly brief exposure durations, this does not affect performance in any of the situations here explored. On the contrary, closely parallel results were obtained at the two exposure durations. As the span, particularly at the briefer exposure duration, falls far short of the capacity of immediate memory, it follows that the effects here demonstrated act at some earlier stage in the recognition process.

One of us (M.J.) was supported by a grant from the Medical Research Council.

REFERENCES

BRICKER, P. D. & CHAPANIS, R. (1953). Do incorrectly perceived tachistoscopic stimuli carry some information? *Psychol. Rev.* **60**, 181–8.

BROADBENT, D. E. & GREGORY, M. (1964). Stimulus set and response set: the alternation of attention. *Quart. J. exp. Psychol.* **16**, 309–17.

CATTELL, J. M. (1886). The inertia of the eye and brain. *Brain*, **8**, 295–312.

CONRAD, R. (1962). Practice, familiarity and reading rate for words and nonsense syllables. *Quart. J. exp. Psychol.* **14**, 71–6.

CONRAD, R. & HULL, A. J. (1964). Information, acoustic confusion and memory span. *Brit. J. Psychol.* **55**, 429–32.

CROSSMAN, E. R. F. W. (1953). Entropy and choice time. The effect of frequency unbalance on choice response time. *Quart. J. exp. Psychol.* **5**, 41–51.

CROSSMAN, E. R. F. W. (1961). Information and serial order in human immediate memory. In *Information Theory, Proceedings of the Fourth London Symposium on Information Theory* (ed. C. Cherry). London: Butterworth.

DAVIS, R., MORAY, N. & TREISMAN, A. M. (1961). Imitative responses and the rate of gain of information. *Quart. J. exp. Psychol.* **13**, 78–90.

FRAISSE, P. & BLANCHETEAU, M. (1962). The influence of the number of alternatives on the perceptual recognition threshold. *Quart. J. exp. Psychol.* **14**, 52–5.

GARNER, W. R. (1962). *Uncertainty and Structure as Psychological Concepts*. New York: Wiley.

GUILFORD, J. P. (1936). The determination of item difficulty when chance success is a factor. *Psychometrika*, **1**, 259–64.

HICK, W. E. (1952). On the rate of gain of information. *Quart. J. exp. Psychol.* **43**, 11–26.

HYMAN, R. (1953). Stimulus information as a determinant of reaction time. *J. exp. Psychol.* **45**, 188–96.

JACOBS, J. (1887). Experiments in 'prehension'. *Mind*, **12**, 73–9.

KINSBOURNE, M. & WARRINGTON, E. K. (1962). The effect of an aftercoming random pattern on the perception of brief visual stimuli. *Quart. J. exp. Psychol.* **14**, 223–34.

KRULEE, G. K., PODELL, J. E. & RONCO, P. G. (1954). Effect of number of alternatives on the visual discrimination of numerals. *J. exp. Psychol.* **48**, 75–80.

MACKWORTH, J. F. (1962). The visual image and the memory trace. *Canad. J. Psychol.* **16**, 55–9.

MACKWORTH, J. F. (1963). The duration of the visual image. *Canad. J. Psychol.* **17**, 62–82.

MILLER, G. A. (1956). The magical number seven, plus or minus two: some limitations on our capacity for processing information. *Psychol. Rev.* **63**, 81–97.

MILLER, G. A., BRUNER, J. S. & POSTMAN, L. (1954). Familiarity of letter sequences and tachistoscopic identification. *J. gen. Psychol.* **50**, 129–39.

SCHMIDTKE, H. (1961). Zur Frage der informationstheoretischer Analyse von Wahlreaktionexperimenter. *Psychol. Forsch.* **26**, 157–78.

SHANNON, C. E. & WEAVER, W. (1949). *The Mathematical Theory of Communication.* Urbana, Ill.: Univ. of Illinois Press.

SPERLING, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.* **74**, no. 11.

TREISMAN, A. M. (1960). Contextual cues in selective listening. *Quart. J. exp. Psychol.* **12**, 242–8.

YNTEMA, D. B. & TRASK, F. P. (1963). Recall as a search process. *J. verb. Learn. verb. Behav.* **2**, 65–74.

(*Manuscript received* 24 *August* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 17–23
Printed in Great Britain

GUIDING DEUTSCH'S MODEL IN REVERSE

BY D. PRITCHATT AND D. H. HOLDING
University of Leeds

In order to test a prediction from Deutsch's 'structural' theory of behaviour twenty-four mice were divided into three groups which were given twenty trials of (*a*) forward trolley guidance, (*b*) backward trolley guidance and (*c*) no trolley guidance, before testing for learning in a four-unit Y-maze. The results showed significant differences between these groups, the forward guidance being the most effective. These results are not consistent with possible predictions from Deutsch's postulates in their present form.

Deutsch (1953, 1960) has proposed a descriptive paradigm for maze learning, with positive reinforcement, which has been used to explain a variety of experiments that have previously been regarded as embarrassing for S-R theories (Deutsch, 1956). However, though some general support for this learning system has been found in other related fields of investigation (e.g. Deutsch, 1964), critical tests in specific maze situations have been rare.

Essentially, the system consists of a series of links down which excitation (motivation) spreads. Each link is associated with a sensory analyser and it is the stimulation of adjacent analysers in temporal contiguity that provides the basis of learning. This consists of the facilitation of excitatory flow between links, rather like the lowering of the resistance of a wire to the passage of electricity. Deutsch describes the process as follows: 'when an animal is put in a maze and it sees cue *a*, which is then followed by cue *b*, then the link whose analyser is set off by cue *a* will be attached to the link whose analyser is set off by cue *b*. If cue *b* is then followed by the stimuli of food, then a similar connexion will take place. As this food will be connected to a primary link, excited by a food deficit, the next time a food deficit occurs the primary link will be excited. This primary link will pass its excitation on to the food link, which will pass it to the cue *b* link, which will pass it on to the cue *a* link. Thus the whole row which has been formed will be excited' (p. 43).

It is not at all clear whether or not the facilitation of excitatory flow between the links is meant to be unidirectional. If it is not, then a curious prediction follows: that the setting up of analysers in reverse sequence, i.e. backward presentation of cues, should be just as effective as the forward sequence during maze learning. Exposure to the sequence *abcde* should permit excitation from *a* to sensitize analysers *b* through *e*, or excitation from *e* to sensitize *d* through *a*. If, on the other hand, the facilitation is unidirectional then no learning should take place with backward presentation. The links would be considered to be joined by 'rectifiers' and transmission from *e* to *a* would not imply transmission from *a* to *e*. The present experiment was therefore undertaken in order to compare forward with backward route learning.

One precaution that must be taken is to ensure that the analysers are identical in the two sequences, otherwise the experiment is invalidated. It is not therefore sufficient to run the animals from goal-box to starting-box since the choice points would stimulate different analysers, depending upon which way the cues were approached. The present experiment attempted to solve this problem by pulling the animal backwards in a trolley from the goal-box with the animal's nose pointing towards the goal-box—the same orientation as that of an animal being pulled forwards to the goal-box; furthermore, the problem of equating feedback information from proprioceptive analysers was solved (this would otherwise have been reversed for the backward group if a free-running backward procedure had been adopted). The procedure seemed feasible since earlier workers have shown that trolley guidance in the maze situation can be effective (Gleitman, 1955; McNamara, Long & Wike, 1956).

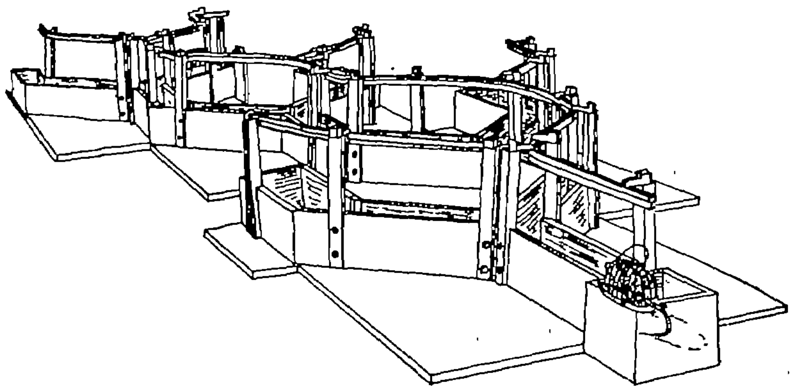


Fig. 1. The four-unit Y-maze. During guidance the trolley cage shown at front right is suspended inside the alley from the overhead rails.

METHOD

Subjects

The experimental animals are twelve male and twelve female mice (*Mus musculus*) approximately 3 months old, equal numbers of both sexes having been selected from two litters of a group of mice that had been cross-bred for ten generations. All animals had received handling but were experimentally naïve. Animals of the same sex and litter were housed together except for the period of food deprivation and testing.

Apparatus

The apparatus consisted of a four-unit Y-maze, each arm of the Y (set at 98° at the narrowest angle) being 9 in. long, 2 in. wide, 3 in. deep. The starting box was 3·5 in. by 4·0 in., and the goal-box was 5·5 in. by 7·5 in. All parts of the maze were open at the top. Plastic curtain flex was fixed 6 in. above the runways, including the blind alleys, so that a trolley could be suspended from it. The trolley, 3·5 in. long and 1·5 in. at the middle where it was widest, was made of 22 s.w.g. nickel wire embedded in a 3 in. plastic base, with a removable rear that could be clipped in (see Figs. 1 and 2). The floor of the maze was lined throughout with irregular patterned washable plastic, and a 100 W bulb suspended centrally over the start-box provided the lighting.

Procedure

The animals were divided into three groups of eight: forward guidance, backward guidance and a control group. Each group contained an equal number of males, females and litter-mates. Half of every group experienced an RLLR maze sequence and the other half an LRRL sequence. All animals were starved prior to the experiment and tested the next day approximately 24 hr later. One animal from each group was tested each day, the order of testing being rotated so that the hours of food deprivation were equalized throughout the groups. A small heap of compressed food pellets (diet 41B) was present in the goal-box throughout the whole experiment for all animals.

Animals in the forward group were put into the trolley and allowed to rest for 5 min in a small cardboard box to enable them to adapt to the trolley since it had been observed previously that most of the bar biting was carried out in the first 5 min period. They were then given twenty pulls through the maze, from start to goal-box, by means of a string attached to a pulley wheel turned by hand at a steady rate, 1 min being allowed from start to goal-box. A pulley wheel was present at each end of the maze throughout the experiment. If an animal turned round in the trolley (which was rare due to the small size of the trolley) pulling stopped and was resumed when the animal turned again. The animal spent 30 sec in the goal-box facing the food reward before being removed and confined in a cardboard box for 2 min between pulls.

The backward animals were treated in exactly the same way except for the fact that they were pulled backwards from goal-box to starting-box with the nose pointing towards the goal-box. The same procedure was adopted for animals in this group which reversed their position in the trolley; these animals were not observed to reverse their heads or bodies more frequently than the forward group.

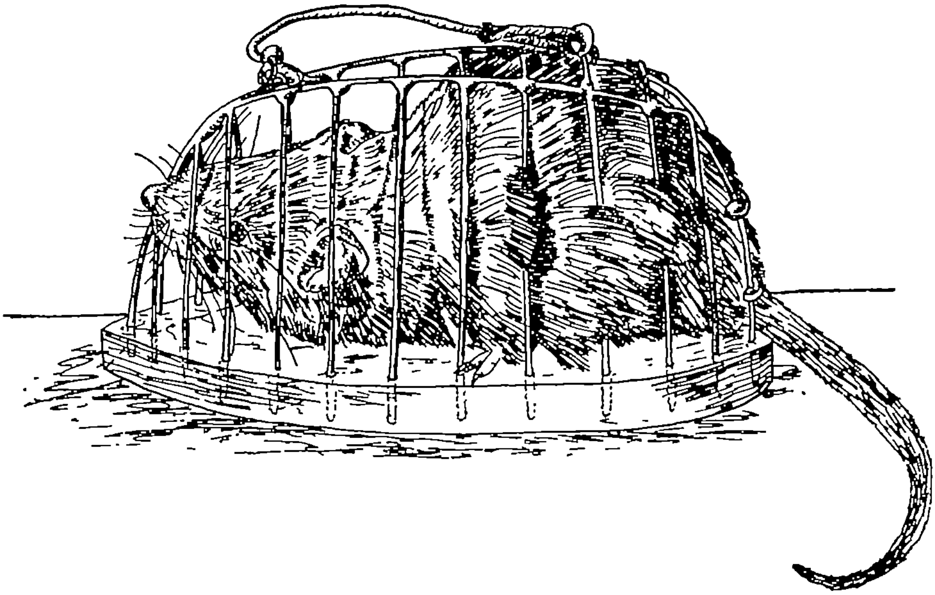


Fig. 2. Trolley cage used in guidance. The dimensions restrict movement to the greatest extent compatible with minimizing emotional reactions to confinement.

After trolley pulls and a 2 min rest, both groups were tested by means of free runs in the maze. If an animal entered a blind, it was taken out and replaced at the starting-box; if it entered the goal-box, it was allowed to eat for 30 sec before being replaced in the cardboard box for 2 min between trials. Testing continued to a criterion of three errorless runs. Between both the free trials and the trolley pulls, the maze was wiped out with a dilute (approximately 0·1 %) solution of disinfectant to remove odours.

Control animals were placed in the cardboard box for 5 min before being tested with free runs in the same manner as the forward and backward groups. Times of running and number of errors were recorded for each of the groups.

RESULTS

The first test run for the control group can be very long and is atypical since the animal has no means of knowing that it will find food in the goal-box; both time and error readings for the first trial were therefore discarded for all groups.

Times

Fig. 3 shows how running times changed with trials in the three groups. The mean times to reach criterion were:

Control—756 sec; Backward—309 sec; Forward—227 sec.

Bartlett's test on the variances of the three groups showed considerable departure from homogeneity ($P < 0{\cdot}001$). The Kruskal–Wallis one-way analysis of variance was therefore used and this test showed an overall difference between times of the three groups ($H = 61{\cdot}9$; d.f. 2; $P < 0{\cdot}01$). Tests of differences between the two gaps using the Mann–Whitney test showed (*a*) a significant difference between Control and Backward ($U = 14$; $P * 0{\cdot}032$); (*b*) no significant difference between Backward and Forward ($U = 22$; $P = 0{\cdot}164$).

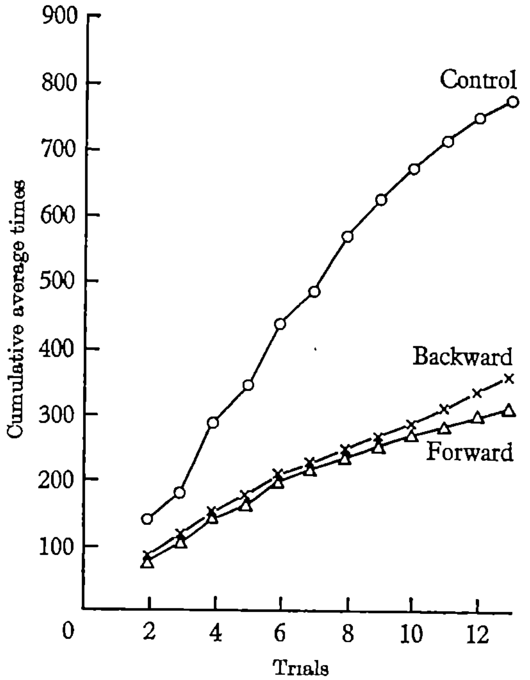


Fig. 3

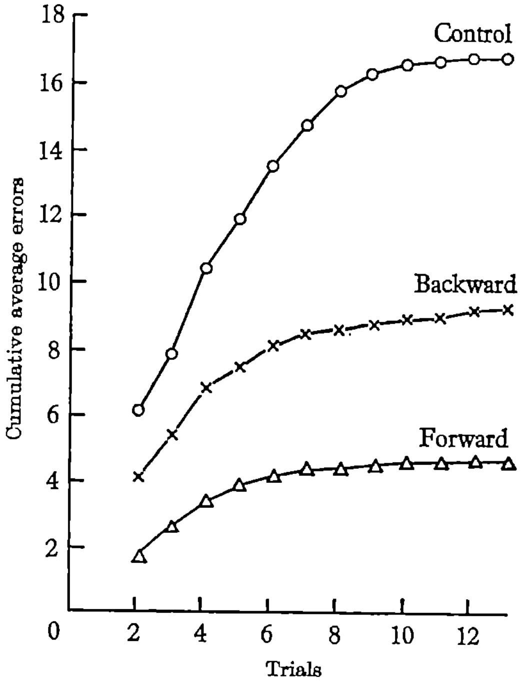


Fig. 4

Fig. 3. Cumulative mean times per trial for groups C, B and F.

Fig. 4. Cumulative errors per trial for groups C, B and F.

Errors

Fig. 4 shows how errors changed with trials in the three groups. The mean errors made in reaching criterion were:

Control—17 errors; Backward—10 errors; Forward—5 errors.

Bartlett's test on the variances of the three groups showed homogeneity of variance. A four-way analysis was therefore carried out, with the results shown in Table 1. The gaps between conditions were tested by means of a range test (Duncan, 1955) and all differences were significant: Control and Forward ($P < 0{\cdot}01$); Control and

Backward ($P < 0{\cdot}05$); Backward and Forward ($P = 0{\cdot}05$). Both time and error measures seem to be quite stable since pilot studies have indicated almost 100% retention when tests are made next day.

Trials

Owing to the use of the replacement technique both errors and errorless runs are confounded within trials. This gross measure is therefore insensitive (Control 10, Backward 9, Forward 7).

Table 1. *Analysis of variance of 'errors'*

Source	S.S.	D.F.	M.S.	F	P
Sequence (RLLR and LRRL)	30·000	1	30·000	1·27	N.S.
Sex	0·042	1	0·042	< 1·00	N.S.
Conditions (Forward, Backward, Control)	592·750	2	296·375	12·56	< 0·001
Litters (L 1 and L 2)	70·042	1	70·042	2·97	N.S.
Residual*	424·791	18	23·599		
Total	1117·625	23			

* Formed by the successive pooling of second- and third-order interactions which failed to reach significance.

DISCUSSION

Though both measures of learning show the same order for the three conditions, the error measure separates the three groups most clearly. This is consistent with the finding of Alonzo (1926), who compared these measures and concluded that guidance 'exerted a slightly greater effect upon errors than upon times in the majority of cases'.

The results show clearly that trolley guidance is effective in the learning of a multiple Y-maze. That some form of guidance can be effective with animals in a maze situation has been shown by Koch (1923), Alonzo (1926), Haslerud (1953), Gleitman (1955) and McNamara *et al.* (1956). These results certainly do not support the generalization made by Maier & Klee (1945) that 'guidance is an effective technique for breaking an old response but not highly effective for teaching a new one'.

In the maze situation there are two things to be learned: (*a*) the expectancy of finding food in the goal-box, and (*b*) the route to goal. With the single T-maze as used by McNamara *et al.* (1956) one is presumably dealing largely with expectancy as stressed by Thorndike (1946). The present experiment is the first to demonstrate the effectiveness of visual guidance, without proprioceptive information, in *route learning* within a maze.

The possible objection that the inferiority of the control group was partially due to their lack of adaptation to the maze situation may be discounted for several reasons. General adaptation is an imprecise term, but may be taken to imply recovery from emotional impediments to performance: in fact, more overt nervousness was observed in the guided animals after their restriction in the trolley cage. Further, it is unlikely that general adaptation would have appreciable effects in a maze as complex as the present one. In any case, reference to Fig. 4 shows that the gap between control and guided animals widens rather than narrows as experience increases.

261296

It is clear, therefore, that backward and forward guidance make a specific contribution to performance in maze learning; but the results are not consistent with either of the predictions that can be made from Deutsch's theory. As it stands, his theory does not postulate unidirectional facilitation between links and therefore predicts no difference between 'Forward' and 'Backward' groups; if unidirectional facilitation were assumed, it would predict no difference between 'Control' and 'Backward' groups.

It appears that the theory is in need of some modification since, though seemingly rigorous, it is curiously inexact in certain respects. Deutsch's own statement regarding the connexion between links is as follows: 'when there is a link whose analyser is stimulated, another link, whose analyser has just ceased to be stimulated, will be attached to the first link, so that this first link will transmit excitation to it' (Deutsch, 1960, p. 43). This does not allow any inference to be made about the ability of the second link to transmit excitation to the first, which might otherwise seem to be implied by the electrical analogy which is introduced later: 'after two links have been connected together, this connexion will have a certain threshold of resistance to the flow of excitation. This threshold, or resistance, will be reduced each time the analysers attached to these two links fire in succession in proportion to the amount of excitation in the links at the time and to the amount of firing from the analysers' (p. 44).

However inexplicit the system may be, its assumptions seem to be clearer when it is applied by Deutsch to specific problems. A critical case in point concerns latent learning in which an animal explores a maze without reward after being introduced into it from different points. Under these conditions analysers will be stimulated in different sequences and, unless bidirectional faciliation occurs between links, the animal would not be able to take advantage of his previous knowledge of the maze to reach introduced food, except when he was retracing a pathway actually taken previously. The layout of fig. 17 (Deutsch, 1960, p. 114) and its description shows that this restriction on performance is not accepted by Deutsch; it appears therefore that bidirectional facilitation is intended.

The present experiment has demonstrated that bidirectional facilitation does occur but that it is not necessarily equal in both directions.* The following amended postulate is therefore proposed: *when there is a link whose analyser is stimulated, another link, whose analyser has just ceased to be stimulated, will be attached to the first link so that resistance to the passage of excitation is reduced between the first link and previously active link and reduced to a lesser extent in the opposite direction.*

---

* It has been brought to the attention of the authors that a hypothesis of this kind has been entertained by Deutsch. It appears that in discussions after his original paper (Deutsch, 1953) he suggested that facilitation of the flow of excitation would be preferentially in the direction that had occurred during learning. However, there seems to be no documentary evidence of his position which would supersede his more recent book (Deutsch, 1960).

REFERENCES

ALONZO, A. S. (1926). The influence of manual guidance upon maze learning. *J. comp. Psychol.* **6**, 143–57.

DEUTSCH, J. A. (1953). A new type of behaviour theory. *Brit. J. Psychol.* **44**, 304–17.

DEUTSCH, J. A. (1956). A theory of insight, reasoning and latent learning. *Brit. J. Psychol.* **47**, 115–25.

DEUTSCH, J. A. (1960). *The Structural Basis of Behaviour*. London: Cambridge University Press.

DEUTSCH, J. A. (1964). Behavioural measurement of the neural refractory period and its application to intracranial self-stimulation. *J. comp. physiol. Psychol.* **58**, 1–9.

DUNCAN, D. B. (1955). Multiple range and multiple *F* tests. *Biometrics*, **11**, 1–42.

GLEITMAN, H. (1955). Place learning without prior performance. *J. comp. physiol. Psychol.* **48**, 77–9.

HASLERUD, G. M. (1953). Anticipative transfer of mechanically guided turns. *J. exp. Psychol.* **45**, 431–6.

KOCH, H. L. (1923). The influence of mechanical guidance upon maze learning. *Psychol. Monogr.* no. 147.

MAIER, N. R. & KLEE, J. B. (1945). Studies of abnormal behaviour in the rat: XVII. Guidance versus trial and error in the alternation of habits and fixations. *J. Psychol.* **19**, 133–63.

MCNAMARA, H. J., LONG, J. B. & WIKE, E. L. (1956). Learning without response under two conditions of external cues. *J. comp. physiol. Psychol.* **49**, 477–80.

THORNDIKE, E. L. (1946). Expectation. *Psychol. Rev.* **53**, 277–81.

(*Manuscript received* 30 *April* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 25–33
Printed in Great Britain

SUTHERLAND'S TWO-STAGE THEORY OF DISCRIMINATION LEARNING AND ITS EXPERIMENTAL SUPPORT

By W. A. MATTHEWS
*Department of Psychology, University of Hull**

Sutherland's two stage theory of discrimination learning is briefly described. Three experiments on overtraining and its effects on reversal and non-reversal shifts which were carried out by Mackintosh (1962, 1963*a*, *b*) and which are held to support this theoretical approach are discussed. The interpretation of these results is criticised in terms of possible differences between the nominal stimulus and the functional stimulus. Mackintosh assumes the identity of these two categories in his interpretations which are claimed to support Sutherland's formulations and to provide difficulties for other theoretical approaches. Control studies of a training and transfer type which might have indicated the effective stimulus were not carried out and, so it is argued, the experiments fail to provide unambiguous support for Sutherland's approach to discrimination learning problems. The identity of the mechanisms mediating transfer along a continuum and the effects of overtraining on reversal and non-reversal shifts is questioned. The lack of compelling evidence to support the notion of sensory dimensional analysers is suggested. Reinterpretation is made in terms of generalizations from a range of learning experiments, and differences between the implications of this approach and that of Sutherland are indicated.

Sutherland (1959) proposed a two-stage theory of discrimination learning for which experimental support has been claimed in papers by Mackintosh (1962, 1963*a*, *b*). The notion is put forward that input from the receptors is submitted to processing by analysers along certain unspecified dimensions. Analyses along only one or a few dimensions occur simultaneously. The analysers produce outputs to which the responses of the animal become attached. The switching in of the analysers is determined by the consequences of the response, both positive and negative consequences having an effect. Presumably positive effects increase the probability that the analyser is switched in and negative effects reduce the probability. Only positive consequences are thought to affect the attachment of the response to the analyser output (Sutherland, 1964).

These views have been applied to two areas of experimental investigation: to the effect of overtraining on reversal and non-reversal shifts and to transfer along a continuum. The overtraining of animals in a two-choice simultaneous discrimination situation has been shown to enable them to reach criterion on a reversal of that task in fewer trials than animals reversed after criterion attainment (e.g. Reid, 1953; Pubols, 1956; Mackintosh, 1962, 1963*a*, *b*). An analysis of the detailed response patterns by Mackintosh (1962) shows that up to the limit of 300 overtraining trials, animals respond to the former S+ for more trials than the criterion group, but then attain criterion on the new S+ more rapidly. A complex environment is thought to be necessary for this behaviour to be shown, but there is no clear evidence as to the necessary environmental conditions. Several experimenters have been unable to produce this effect in a position discrimination task in a T-maze (e.g. Hill, Spear & Clayton, 1962; D'Amato & Jagoda, 1962; Clayton, 1963). In the last investigation an attempt was made to manipulate irrelevant intra- and extra-maze cues, but in two experiments no effect was shown.

---

* Now at Department of Psychology, University of Southampton.

The 'transfer along a continuum' phenomenon was first shown clearly with rats in a jumping-type situation by Lawrence (1952), who used a series of grey cards. The task was to learn a discrimination between two similar greys. There were three comparable groups of animals: one trained for 80 trials on the difficult discrimination: one trained for 30 trials on an easier pair of greys and then given 50 trials on the difficult discrimination; and one trained for the first 30 trials on three different problems (10 trials each) before being given 50 trials on the difficult discrimination. The latter two groups attained criterion 20 trials earlier than the first group and showed fewer errors at each stage of the experiment. This has also been observed in tropical fish (Hemmings, unpublished), where pretraining on rectangles approximating in a stepwise fashion to a square led to later efficient performance on a circle/square discrimination, which tropical fish do not easily solve (Hemmings & Matthews, 1963; Matthews, 1964). It has also been reported in octopuses (Sutherland, Mackintosh & Mackintosh, 1963).

To explain the overtraining and reversal effect, it is assumed that overtrained animals have continued to switch in the appropriate analyser during the overtraining period so that when reversal occurs the animal continues to switch in the same analyser for a longer period than the criterion group, even though reward is not now following most responses. When the response attachment to the now incorrect output from the appropriate analyser is extinguished, the animal will attach the response to the other output from the same analyser and this stage only needs to be learnt on reversal. This implies a long series of responses to the incorrect stimulus and a fairly short relearning period. For the non-overtrained group, the appropriate analyser is assumed to be switched out as a result of the non-reinforced trials and inappropriate analysers emit outputs to which response attachment will be tried. As these will not be consistently followed by reward, the animal will continue to sample until it resumes using the appropriate analyser but now begins to respond to the other output. Thus the new criterion will be attained more slowly by the group initially trained to criterion. If the environment lacks irrelevant cues, no differences between overtrained and criterion trained groups should occur on reversal. If irrelevant cues are introduced on reversal, this should not affect the effects of overtraining, as the appropriate analyser is thought to be switched in.

Mackintosh has reported a series of experimental results which are claimed to support this approach and to create difficulties for other theoretical proposals, e.g. Harlow's error factor theory (Harlow, 1959) and Pubols (1956). An assessment of these results follows.

Mackintosh's experiments

Mackintosh (1963*a*) reported an experiment in which the effect of irrelevant cues, the place of their introduction and the subsequent effects on reversal were investigated. Two groups of rats pretrained with black and white squares as the discriminanda were reversed on either black and white squares or black and white rectangles. A further group was trained and reversed on black and white rectangles. The orientation of the rectangles varied randomly in each case. Within each of these groups were subgroups given either 0 or 150 trials beyond criterion. The animals were run in a modified Lashley jumping stand and jumped to platforms below each window. The stimulus was fixed in the middle of the window with a distance of about 4·5 cm from the platform to the base of the shape. There were no significant differences between the three main groups in trials to criterion on the initial discrimination, but the group trained on the rectangle took slightly longer. The results indicated that overtraining had the expected effect on the reversal scores in each case, with small differences between these groups in trials to criterion on the new task. The group trained and reversed on black and white squares showed the least difference between the criterion trained and overtrained groups, and this could be ascribed to the fewer irrelevant cues in this discrimination. The two remaining groups which had rectangles varying in orientation in the discriminandum showed larger differences between the criterion and overtrained sections of each group.

The differences between the groups trained and reversed on the same stimuli are compatible with both Sutherland's model and Harlow's error factor theory, but the results of the group trained on squares and reversed on rectangles is thought to be not compatible with error factor theory but to fit the predictions of Sutherland's model. This group showed no significant differences in the reversal scores from that of the rectangle trained and reversed group. As incorrect response tendencies to the rectangles could not have been adapted out by this group in training, more inappropriate responses on reversal are predicted by error factor theory and this should slow down reversal criterion attainment. No differences are predicted by the analyser model. Although Mackintosh claimed that the results support only Sutherland's model, they are not necessarily incompatible with error factor theory for the interpretation made by Mackintosh makes one critical but undemonstrated assumption: that the nominal stimuli were the functional ones, and that the logical breakdown of the stimulus into brightness, shape and orientation was what the rats saw and used. This is not necessarily the case (e.g. Dodwell, 1961; Sutherland, 1961). It has been shown that rats learn something about constant irrelevant cues (Jeeves & North, 1956; Bitterman & Coate, 1950) and we have found this clearly in this department using pigeons (Williams, unpublished). In the experiment above, the animals may have learned something about a black or white flat-based right-angled shape, possibly limited to the base of the figure. The initial slightly slower learning of the rectangle-trained group may be due to the varied locus on the door of the stimulus which changed in orientation. On reversal, this varied locus may still be affecting the two non-overtrained subgroups reversed on the rectangle but is not affecting the animals reversed on the square, so accounting for the fewer trials to criterion on the reversal of the non-overtrained square group. For the overtrained group, the continued reinforcement may lead to an increased discrimination of what is, for the animal, the relevant cue. Thus, locus of the shape and some other aspects of the stimulus come to be disregarded as each animal in each group develops its own effective stimulus. On reversal, no difference is seen between the two overtrained groups which were reversed on rectangles because the ostensible irrelevant cue is not in fact seen as one. The predictions of a hypothetical experiment might serve to distinguish this notion of an increased discrimination of the relevant cue from the selection of appropriate analysers (Sutherland, 1959).

Assume a simple discrimination training situation in which the stimuli are one black and one white square, with animals overtrained on the task. After overtraining, the animals are given transfer tests to black and white rectangles, triangles and circles of equal area (if rats or fish are not being used). An analyser approach would predict equally efficient transfer to all of those shapes, because analysis in terms of shape would have ceased in the training trials as it would fail to give differential outputs, and analysis in terms of brightness only would have occurred. On the approach proposed here, initial analysis might have been in terms of brightness and contour. Thus on transfer, different degrees of responding would be predicted and these would be correlated with the extent to which the transfer shapes differed from the training shapes (as used by each animal). The increased discrimination of the relevant cue could occur by a normal S–R continuity process or by processes visualized in error factor theory, where the actual location of the animal's cue is adapted out as an error factor. This should occur in every response situation where, in addition to other discrimination processes, the orienting behaviour of the animal leads to variability in the precise location of the stimulus. The experiment by Mackintosh (1963*a*) did not include controls adequate to validate the essential point of Mackintosh's interpretation that brightness, shape and orientation were treated as separate stimulus elements.

Mackintosh (1962) made a further attack on other theoretical approaches in a report that animals which are overtrained on a simultaneous discrimination perform better than a criterion group on a reversal shift but worse than a criterion group on a non-reversal shift. Facilitation or interference increases as the degree of overtraining increases. This, it was claimed, is because the correct analyser has been switched in for the reversal shift, but an incorrect one for the non-reversal shift; the firmness of switching in increases as the number of overtraining trials increases. Experimentally, rats jumped to a black or white 12 cm square attached to the centre of a 15 cm brown hardboard background, with reward for a correct choice. The reversal shift was a brightness one. The non-reversal shift was a response to the same background on which had been glued three 8 × 2 × 0·5 cm blocks of wood painted mid-grey. Between each bar was a 1 cm gap. The blocks were oriented either vertically or horizontally. Ostensibly this was a change from a brightness to an orientation discrimination (in the non-reversal shift). It is far from clear what the animals were in fact responding to, but it is likely that they learned something about a black or white straight-lined right-angled shape. On the non-reversal shift, the apparent switch to orientation may be one to grey shapes with similar characteristics to the training shape, but which have been differently affected by the amount of training, i.e. this was not in fact a non-reversal shift. If training leads to an increasing discrimination of the relevant stimulus array, owing to the presentation of both positive and negative instances, it seems not unreasonable to hypothesize that this may continue in overtraining even though the response to the negative stimulus is infrequent. Thus the greater the extent of overtraining, the more limited may the stimulus aspects become to which response is made. When the non-reversal shift is made, direct transfer from the training to the testing shape will be less, as the degree of overtraining delimits the functional stimulus. Thus slower criterion attainment as the degree of overtraining increases would be predicted on the non-reversal shift. Mackintosh claimed to have controlled for direct transfer from training to reversal, but this was only on brightness differences to horizontal and vertical striations, and the interaction term was non-significant in the analysis of variance that he carried out on the results. On the argument proposed here, the appropriate transfer term would be the main effect of degree of overtraining which was significant. ($F = 16{\cdot}89$, $P < 0{\cdot}001$). It is also interesting that a significant effect was found for the horizontal orientation in the non-reversal shift problem. This could be due to direct transfer from a straight-line-based figure as a probable subjective stimulus in training and/or to the well-known preference of rats for horizontal over vertical striations.

In the reversal shift, it is proposed that if the stimulus were a black or white shape, the facilitation correlated with the degree of overtraining may be due to three factors. The first is that responses to non-rewarded shapes have negative effects on the animal leading to avoidance behaviour (e.g. Amsel, 1962) and that these effects dissipate during overtraining. The second is that as the effective stimulus becomes more limited as overtraining proceeds, the generalization of extinction during the first reversal trials should be reduced as overtraining increases. Thus there is a reduced likelihood of all the aspects of the new subjective stimulus being affected by the extinction processes early in the reversal trials; and thirdly, it is likely that this stimulus will contain components of the previously positive stimulus to which an approach response has not been completely extinguished. Subsequent attainment of criterion should therefore be speeded. Again, it would be argued that in this experiment, the required controls to look at the functional stimulus were not run.

A further paper by Mackintosh (1963*b*) reports two additional experiments which must be considered. The first concerns the course of extinction of a simultaneous discrimination (learned to criterion or to 150 trials beyond it), and repeats an earlier finding that the speed of attainment of the new task criterion is located in the learning phase for the overtrained group. This group took a greater number of trials to reach extinction criterion than did the criterion trained group. A subsidiary finding and its interpretation is of interest here. Latency is used as a measure of extinction, and although starting from similar points after training, the response latency of the overtrained group to the former S + rose more rapidly than that of the criterion group over the first 40 trials, but levelled off at a lower point than that of the criterion group. On the 10 extinction criterion trials, however (5 responses to S + and 5 responses to S −), the responses of the criterion group to both S + and S − occurred at a similar speed. The responses of the overtrained group were faster to the S + than to the S −. This, it is claimed, is because the appropriate analyser is switched in even on the extinction criterion trials, and the overtrained animal is still discriminating between the stimuli, but the criterion group are not now discriminating. This assumption seems unwarranted. If the criterion group are sampling other stimulus sources, it is difficult to see why the selection of these should be slower than that of a partially extinguished response attachment to a partially extinguished analyser. It is also difficult for Mackintosh to explain the more rapid increase in latencies for the overtrained group over the first forty extinction trials in similar terms, as presumably the analyser used and the response attachment to the output should have been more firmly made for this group than for the criterion group.

Experiment 2 of this paper is an example of the behaviour of rats in a discrimination situation for which the analyser model is claimed to be particularly appropriate and it is thought to provide difficulties for continuity theory. It is a repeat of a three-stage experiment by Goodwin & Lawrence (1955) with the addition of an overtrained group. In stage 1, all animals learned a rectangular brightness discrimination, with orientation (either horizontal or vertical) irrelevant. In stage 2, all subjects learned a rectangle orientation discrimination for 150 trials (with brightness irrelevant). In stage 3, half of the subjects relearned the original discrimination, and half learned its reversal. Group A were trained to criterion in stage 1. Group B had 150 trials beyond criterion. Criterion was attained in approximately the same number of trials by each group. In stage 2, claimed to be a non-reversal shift, group B attained criterion in significantly more trials than group A. As both groups had 150 training trials in stage 2, it must be noted that group A had 50 overtraining trials in stage 2, whereas group B had only 17. In stage 3, group B attained criterion on the initial discrimination faster than group A (63 *versus* 72 trials) and were also faster on the reversal (90 *versus* 173). This is thought to show that the differential habit strength to the S+ and S− of stage 1 are left largely unaffected by extinction, but that this produces a greater equalizing effect for group B, which is implied by the stage 3 superiority of group B. It is also noted that group B learned stage 2 more slowly than did group A. Again, problems of the effective stimulus must be raised. If the stage 1 learning was in terms of black or white shapes, then generalization to the stage 2 problem could occur in terms of shape components in the orientation task. However, this direct transfer should be less for group B, if discrimination learning and a reduction in the effective stimulus had continued in overtraining. Additionally more responses should have been made to the stage 1 effective stimulus in stage 2 by group B, as this group had a greater number of reinforcements in stage 1, further slowing down the attainment of criterion on the new task. The difference in the number of trials to criterion in the original discrimination task at stage 3 between groups A and B can be accounted for by the simple effect of overtraining in stage 2 for group A on the ostensible orientation discrimination. The difference between groups A and B in trials to criterion on the reversal shift may be due partly to the stage 2 overtraining for group A, and partly to the fact that group B had 150 more trials interpolated between attaining criterion on stage 1 and reversing on stage 3 than had group A. This could have resulted in the effects of non-reinforced responses to S− of stage 1 (which was S+ in the reversal shift of stage 3) being less noxious for group B than group A. Thus these findings which have been claimed to be difficult for continuity theory, and to have provided strong support for Sutherland's analyser model, appear to be less important than has been claimed.

Other empirical findings

A consideration of some empirical findings in other fields of learning suggests another way of approaching the data in this area. The implications of an integration of these findings are in certain ways different from those derived from Sutherland's model, even though mechanisms are not suggested. The following five points seem relevant:

1. The effects of non-reward are not neutral but to some degree noxious. The probability of response to a non-noxious stimulus will be greater than the probability of response to a noxious stimulus.
2. The noxious effects of non-reward become less strong as time passes or other activity intervenes between the initial experience and later presentation of the same stimulus.

3. Discrimination improves as a function of the presentation of positive and negative instances with correlated reward and non-reward. After this process has begun, it is hypothesized that discrimination may continue to improve with less and less frequent presentation of the S–. Empirical evidence does not exist to support this last notion, but could be obtained.
4. The more frequently has a response to a stimulus been reinforced during discrimination training, the more persistent will that response be in the absence of reinforcement. This applies up to certain limits of overtraining only (D'Amato, Schiff & Jagoda 1962).
5. Transfer of responses only occurs to stimuli having similar components.

The following effects should occur in overtrained animals compared with animals trained to criterion. First, the negative after-effects of non-reward should be reduced, and this should be correlated with a reduced tendency to avoid the previously S –. Secondly, there should be greater persistence in the responses made to the S+ in training if reward is withdrawn (within certain limits of overtraining). Thirdly, the aspects of the stimulus to which the animal responds should be reduced as training proceeds, leading to reduced generalization of extinction. Thus a greater probability of responding to the new S+ should occur in the overtrained group. This fits well with the results reported by Mackintosh (1962, 1963*a*, *b*).

It is possible to cope with these experimental results in terms of a continuity theory with several components. Certain distinctions may be made between this approach and that of Sutherland.

1. The extent of facilitation or its reverse is limited by the objective stimulus similarities in the response situation (see the earlier discussion of Mackintosh, 1962). Thus, overtraining on a circle/ellipse discrimination should not affect later discrimination of a square/triangle discrimination. In Sutherland's terms there seem to be two types of non-reversal shift, intra- and extra-dimensional. Thus in the example above, overtraining on a circular shape should facilitate performance on a shape without circular components, as compared to a criterion-group's performance. This assumes a 'shape' analyser. If, however, the groups pretrained on circular components were switched to amorphous masses of colour as the stimuli to be discriminated, Sutherland would predict that the overtrained groups should be hindered in criterion attainment on the new task.

2. Experience given to criterion-trained groups (*a*) of presenting the initial S – but not allowing a response to it, in an interval between criterion attainment and reversal, or (*b*) of lengthening the period between criterion attainment and reversal may reduce the size of any difference in trials to criterion on reversal between the criterion and overtrained groups. These experiences should not alter the overtraining effect according to the analyser model.

3. This approach would predict that the overtraining effect on reversal would be reduced for groups of animals which have increasingly noxious stimulation correlated with the S – in initial training.

4. There should be no necessary relationship between showing the overtraining effect and showing what has been labelled 'transfer along a continuum' in any one species.

These two phenomena have been analysed by Sutherland in terms of the analyser model. Transfer along a continuum (that is, the facilitation on a difficult discrimination by prior training on an easier discrimination) is ascribed to the pretrained animals switching in the appropriate analyser more reliably in pretraining and therefore transferring this strategy to the difficult discrimination. Difficulty only arises now in attaching the appropriate response to the similar analyser outputs. The animals trained on the difficult discrimination from the beginning are thought to vary analysers too, due to the difficulties of correlating responses and reward with the analyser outputs. Three experiments which purport to show this have been mentioned earlier. One of these (Hemmings, unpublished) has reported this phenomenon in tropical fish. Sutherland comments that failure to find both of these phenomena in one species would be evidence against the theory. Matthews (unpublished) has failed to find the overtraining and reversal effect and the related effect of increasingly rapid criterion attainment as the number of reversals increase in tropical fish. However, this is weak evidence against the theory for it is clear that failure to find a discrimination in one situation may not be evidence of a deficit in the animal in all situations (cf. Munn, 1930; Lashley, 1938).

More substantial grounds for criticism of the 'transfer along a continuum' notion exist. In the Sutherland *et al.* (1963) experiment, the control group was trained on a square and a similar parallelogram. The experimental group was given prior training on four different parallelograms paired with the square, before the most difficult discrimination was presented for 3 days. On these days the experimental group got 73 % correct and the control group got 54 % correct, after the same total number of days of training; but this seems little different from a normal transfer experiment using similar components as the basis for transfer, with more positive transfer being shown in the experimental group which had attained approximately 79 % correct choices at each stage of pretraining, as compared with approximately 50 % correct at similar training stages for the control group. A more reasonable test of this hypothesis would be available if each group attained a similar but low degree of learning at each training stage, and then the performance of the experimental group was found to be more efficient than that of the control group on the final difficult discrimination. What seems to be required for transfer along a continuum to be shown to be a different phenomenon from simple transfer via similar components is that pretraining be given to a similar criterion for the experimental and control groups. The pretraining for the experimental group should be given on stimuli which are thought to be along a dimension, but from which little generalization could be expected to the critical pair. One example of this might be to pretrain on colours at the distant end of the spectrum from that to which the critical pair to be discriminated belong. Facilitation for the experimental group would be strong evidence for a physical dimension analyser, and evidence against the proposal that all that has been shown by the three experiments mentioned earlier is facilitation via similar components after a more efficient training procedure for the experimental group.

These criticisms of the experimental support for Sutherland's model have emphasized ambiguities in interpretation that are possible owing to the design of the experiments. Sutherland (1961) made the point that the functional stimulus may not be the nominal one in his summary of experiments on shape perception; and this has been emphasized by the rather unsuccessful attempts to formulate theories of shape perception to date. Similarly, analyser constructs cannot be used precisely in experimental situations unless some attempt is made to specify the functional stimulus.

Mackintosh's interpretations of these experiments have hinged on this undemonstrated assumption and so have failed to produce any strong evidence in favour of or against any particular theoretical approach to discrimination learning.

I wish to acknowledge valuable discussions with C. I. Howarth and D. I. Williams.

REFERENCES

AMSEL, A. (1962). Frustrative non reward in partial reinforcement and discrimination learning. *Psychol. Rev.* **69**, 306–26.

BITTERMAN, M. E. & COATE, W. B. (1950). Some new experiments on the nature of discrimination learning in the rat. *J. comp. physiol. Psychol.* **43**, 198–210.

CLAYTON, K. N. (1963). Reversal performance by rats following overlearning with and without irrelevant stimuli. *J. exp. Psychol.* **66**, 255–9.

D'AMATO, M. R. & JAGODA, H. (1962). Overlearning and position reversal. *J. exp. Psychol.* **64**, 117–22.

D'AMATO, M. R., SCHIFF, D. & JAGODA H. (1962). Resistance to extinction after varying amounts of discriminative or non discriminative instrumental training. *J. exp. Psychol.* **64**, 526–32.

DODWELL, P. C. (1961). Coding and learning in shape discrimination. *Psychol. Rev.* **68**, 373–82.

GOODWIN, W. R. & LAWRENCE, D. H. (1955). The functional independence of two discrimination habits associated with a constant stimulus situation. *J. comp. physiol. Psychol.* **48**, 437–43.

HARLOW, H. F. (1959). Learning set and error factor theory. In *Psychology: A Study of a Science*, study 1, vol. 2, pp. 492–537. Ed. S. Koch. New York: McGraw Hill.

HEMMINGS, G. & MATTHEWS, W. A. (1963). Shape discrimination in tropical fish. *Quart. J. exp. Psychol.* **15**, 273–8.

HILL, W. F., SPEAR, N. E. & CLAYTON, K. N. (1962). T-maze reversal learning after several different overtraining procedures. *J. exp. Psychol.* **64**, 533–40.

JEEVES, M. A. & NORTH, A. J. (1956). Irrelevant or partially correlated stimuli in discrimination learning. *J. exp. Psychol.* **52**, 90–4.

LASHLEY, K. S. (1938). The mechanism of vision. XV. Preliminary studies of the rat's capacity for detail vision. *J. gen. Psychol.* **18**, 123–93.

LAWRENCE, D. H. (1952). The transfer of discrimination along a continuum. *J. comp. physiol. Psychol.* **45**, 511–16.

MACKINTOSH, N. J. (1962). The effects of overtraining on a reversal and non reversal shift. *J. comp. physiol. Psychol.* **55**, 555–9.

MACKINTOSH, N. J. (1963*a*). The effect of irrelevant cues on reversal learning in the rat. *Brit. J. Psychol.* **54**, 127–34.

MACKINTOSH, N. J. (1963*b*). Extinction of a discrimination habit as a function of overtraining. *J. comp. physiol. Psychol.* **56**, 842–7.

MATTHEWS, W. A. (1964). Shape discrimination in tropical fish. *Anim. Behav.* **12**, 111–15.

MUNN, N. L. (1930). Visual pattern discrimination in the white rat. *J. comp. Psychol.* **10**, 145–66.

PUBOLS, B. H. JR. (1956). The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.* **49**, 243–8.

REID, L. S. (1953). The development of non continuity behaviour through continuity learning. *J. exp. Psychol.* **46**, 107–12.

SUTHERLAND, N. S. (1959). Stimulus analysing mechanisms. In *Proceedings of a Symposium on the Mechanization of Thought Processes*, vol. 2, pp. 575–609. London: Her Majesty's Stationery Office.

SUTHERLAND, N. S. (1961). The methods and findings of experiments on the visual discrimination of shape by animals. *Exp. Psychol. Soc. Monogr.* no. 1.

SUTHERLAND, N. S. (1964). Visual discrimination in animals. *Brit. med. Bull.* **20**, no. 1, pp. 54–9.

SUTHERLAND, N. S., MACKINTOSH, N. J. & MACKINTOSH, J. (1963). Simultaneous discrimination training of octopus and transfer of discrimination along a continuum *J. comp. physiol. Psychol.* **56**, 150–6.

(*Manuscript received* 19 *November* 1964; *revised manuscript received* 6 *September* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 35–43
Printed in Great Britain

POSITIVE AND NEGATIVE RECENCY IN TWO-CHOICE GUESSING

By H. C. A. DALE

Medical Research Council, Applied Psychology Research Unit, Cambridge

According to the nature of the run of stimuli and outcomes, either positive or negative recency may be exhibited in the initial responses to a two-choice guessing task. When red and black two-digit numbers printed in two columns were used as stimuli, subjects responded to a run on one colour with positive recency, but to a run on one column with negative recency. Responses were found to be related to prior expectations in a way which occurs with a rational Bayesian decision process. With a *very* long run on one position (either column) the marked negative recency which occurred initially eventually gave way to positive recency. Three experiments are reported.

In an earlier study by the present writer (Dale, 1965) factors which affect the initial reaction to runs in two-choice guessing tasks were explored. The experiments were prompted by an observation made by Green (1958). Green employed a novel guessing task in which subjects were given a 'clue' adjective and had to guess a 'target' adjective related to the clue. If a run developed in which the targets were all synonymous with the clues or were all antonyms of the clues, subjects showed positive recency. This is very different from the more usual negative recency effect found in two-choice guessing, when, for example, a run of 'heads' in coin guessing induces a strong tendency to predict 'tails'.

Green's adjective guessing task differed in a number of respects from tasks known to yield negative recency. In addition to the use of clues: (*a*) subjects had to generate the answers, (*b*) the stimuli varied from trial to trial, and (*c*) the subjects were not specifically instructed that a binary choice was to be made. Separate manipulation of these factors in the earlier study indicated, however, that the principle influence upon recency was the use of the clue–target relation in place of the simple choice between alternative stimuli (cf. Dale, 1965). The present investigation began by testing the generality of this distinction with new stimulus material.

(The subjects used in all three parts of this study were young newly enlisted men. Each subject participated in only one condition.)

Experiment I

Numerical material was substituted for the adjective guessing task used previously. The clues were 3-digit numbers between 100 and 199, while the targets were pairs of 2-digit numbers derived from the clues by dropping the first digit. One number of each pair was similar to the clue in that the remaining two digits were in the same order. This corresponded to the synonym. The other was opposite, having the order of the digits reversed, which corresponded to the antonym.

Method

Stimulus material

The material is shown in Table 1. All targets in the left-hand column were similar to the clues while all those in the right-hand column were opposite.

Procedure

Two conditions, *A* and *B*, were employed. Clues were used in condition *A* but not in *B*. The procedure for condition *A* is described in the printed instructions which were given to the subject: 'This is a sort of guessing game. I have a list of numbers between 100 and 200 which I am going to read out one at a time, Each of these provides a clue to a target which is a number between 0 and 100. The first target will be taken from the top row of numbers on this page, i.e. it will be either 34 or 43. The second will be one of the second row, and so on. Your task is to guess each time what the target number will be. Write down each clue, as I give it, in the appropriate space. Then guess which of the two numbers listed here is the target. Mark each number you guess by putting a ring round it. As soon as you have made each guess, put your pencil down and look up. When I see that you have all finished I will read out the answer. You will then go on to the next pair. Always make a guess even though you may not be sure which it will be.'

Table 1. *Clues and targets in Expt. I*

Item	Clues	Targets		Item	Clues	Targets	
1	134	34	43	11	118	18	81
2	173	73	37	12	162	62	26
3	116	16	61	13	128	28	82
4	147	47	74	14	194	94	49
5	148	48	84	15	176	76	67
6	145	45	54	16	135	35	53
7	132	32	23	17	121	21	12
8	194	94	49	18	117	17	71
9	163	63	36	19	165	65	56
10	141	41	14				

In condition *B* the same target numbers served as alternatives and on each trial the subjects had to guess which number had been nominated by the experimenter as correct. No clues were given. The first part of the instructions was changed accordingly to read: 'This is a sort of guessing game. I have a list of numbers which I am going to read out one at a time. The first will be taken from the top row of numbers on this page, i.e. it will be either 34 or 43. The second will be one of the second row, and so on. Your task is to guess which number from each row is on my list. Mark each number you guess by putting a ring round it. Only mark one at a time.'

In condition *A* the relevant feature of the target numbers was their correspondence with the clues. The sequence of correct answers was 1 1 1 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 for the nineteen trials, where '1' refers to a target of one kind, and '0' to one of the other kind. In condition *B*, the relevant feature was the position, i.e. whether the number was chosen from the left-hand or the right-hand column. Since in *A* targets with digits in the same order as the last two digits of the clue were always in the left-hand column, the actual targets correct in condition *A* were precisely the same as those correct in condition *B*. The fourth digit in the run structure was critical; if a subject exhibited positive recency he would be incorrect on this trial, but if negative recency occurred he would be correct. Thus the proportion of correct answers to the fourth trial is an index of recency.

Twenty-six subjects tested in two batches were given condition *A*; thirty-three subjects, also tested in two batches, received condition *B*. The selection of targets was counter-balanced to control for bias towards one kind of target, so that for half the subjects the first correct target was similar to the clue or came from the left-hand column, while for the remainder it was dissimilar, or came from the right-hand column.

Results and discussion

The proportion of subjects who predicted correctly on each trial is given in Table 2 below. It can be seen from the response to trial 4 that in condition *A* 96 % of the subjects predicted the run would continue. In condition *B*, by contrast, only 15 %

predicted the run would continue. Negative recency in this latter condition had developed rapidly and was evident by trial 3. The difference between conditions in the numbers correct on trial 4 is statistically highly significant ($\chi^2 = 34{\cdot}9$, D.F. = 1, $P < 0{\cdot}001$). Both proportions differ significantly from the chance level of 50% ($P < 0{\cdot}01$ in each case). A similar but less-marked difference appeared on trial 8 when the second reversal in the stimulus series occurred ($\chi^2 = 4{\cdot}2$, D.F. = 1; $P < 0{\cdot}05$).

The phenomena obtained with adjectives in the earlier study were thus substantiated by the present experiment in which different materials were used. This indicates that there is some general validity to the conclusion that the use of clues facilitates positive recency, while guessing the source is associated with negative recency.

A slight modification of the instructions had incidentally eliminated a possible artifact. In the previous study, when clues were used, the subjects had been told that the target was 'related to the clue in some way'; this they might have slightly misconstrued as 'related to the clue in one way' and they might have been predisposed to seek a simple rule as a consequence. By contrast, when clues were absent, the instructions had stated that the experimenter's list was made up of items from the two lists of alternatives. This implies that items from both lists would be included, so that when a run from one list appeared it would be reasonable to assume that an item from the alternative list was due. It will be noted that the instructions for the present experiment were carefully reworded to avoid these biases.

EXPERIMENT II

The many two-choice guessing situations which have been shown to yield negative recency are characterized by their barrenness. Thus Jarvik's subjects had to anticipate whether he would call out 'check' or 'plus' (Jarvik, 1951) and Nicks's subjects had to predict whether a red or a green lamp would be lit (Nicks, 1959). In both their experiments the only difference between the stimuli was that relevant to the choice. It is pertinent to ask whether negative recency also occurs where the stimuli vary in more than the one relevant dimension.

Expt. I and the experiments in the previous study (Dale, 1965) provide a partial answer to this question. The target adjectives varied from trial to trial, as did the particular 2-digit numbers employed. With the numbers, the subjects could have attempted to relate the correct outcomes to the relative size of the numbers instead of to the list from which they were drawn, but this secondary feature of the stimuli was not very prominently displayed. In the present experiment the material was modified so that the subjects were presented with two prominent ways of categorizing the alternatives. The numbers were again used, but this time one selected at random from each pair was printed in red ink. Under one condition subjects were exposed to regular choice from one position, colour being irrelevant. Under the other condition colour was critical and position irrelevant.

Method

Stimulus materials and sequences

Target lists and instructions were basically the same as in condition *B* of Expt. I. Two sequences were prepared which differed in the allocation of colour to position. They are shown in Table 2, condition *D*, where the proportions are given in two weights of type; one weight corresponding to one relation between colour and position and the other weight to the other relation. For counter-balancing there were two versions of each sequence: in version *a* the trials italicized are those where the red number was in the left-hand column; in version *b* the relation was reversed. It is to be noted that colour and position were confounded for the first three trials in sequence 1, whereas colour varied independently of position on the second trial in sequence 2.

Procedure

The two experimental conditions were condition *C*, in which the targets were all selected from the same column, and condition *D*, in which the targets were all of the same colour. Eighty subjects were tested under each condition; none of the 160 subjects had taken part in Expt. I,

Under condition *C* there were two groups of subjects: for one all targets were on the left, for the other all targets were on the right. Each batch of subjects was split into four subgroups so that all four sets of stimulus material could be used simultaneously.

Under condition *D* there were four groups, one with each set of material. Each batch of subjects was split, half having version *a* lists and half version *b* lists, so that for half of each batch all the targets were red, while for the other half they were all black.

At the end of the session the subjects were asked to give a brief written description of the way they had made their predictions.

Results and discussion

It is possible to examine the reaction to the runs as a whole because the rule was applied consistently throughout the complete run of nineteen trials in each condition. The tabulated results reveal a distinct difference between the two conditions (cf. Table 2). When the targets were all of one colour a simple acquisition curve was obtained, the initial rate varying according to the sequence. When the targets were all taken from the same position (in either the left- or right-hand column) a complex, three-phase response was obtained: (i) from trial 1 to trial 2 there was a slight increase in the number choosing correctly; (ii) from trial 2 to trial 4 there was a marked decrease; then (iii) from trial 5 onwards the number correct increased steadily, ultimately reaching an asymptotic level of nearly 90 %.

Table 2. *Proportion of subjects correct on each trial in Expts. I and II*

	Condition				
	A Numbers, clues	*B* Numbers, no clues	*C* Constant position	*D*, Constant colour	
				Sequence 1	Sequence 2
n ...	26	33	80	40	40
Trial 1	0·62	0·61	0·42	0·40	*0·50*
2	0·85	0·58	0·44	0·47	0·42
3	0·88	0·27	0·30	0·40	0·77
4	0·04	0·85	0·20	*0·85*	*0·77*
5	0·65	0·61	0·32	*0·82*	*0·65*
6	0·81	0·45	0·42	0·75	*0·57*
7	0·77	0·52	0·36	*0·77*	0·85
8	0·38	0·73	0·56	0·80	0·72
9	0·31	0·58	0·61	*0·90*	*0·80*
10	0·43	0·36	0·76	*0·82*	0·80
11	0·38	0·39	0·67	*0·85*	0·70
12	0·46	0·39	0·82	*0·72*	0·67
13	0·50	0·39	0·80	0·90	0·72
14	0·58	0·27	0·82	0·82	*0·92*
15	0·54	0·52	0·86	*0·77*	0·80
16	0·77	0·39	0·87	0·77	*0·85*
17	0·62	0·39	0·86	0·82	*0·75*
18	0·58	0·39	0·90	0·80	*0·82*
19	0·65	0·45	0·80	0·80	*0·70*

Recency effects in the previous experiments were gauged by the response on trial 4 which followed a run of three similar outcomes. The same criterion can be applied again. After the run of one colour on trials 1–3, sixty-five out of eighty subjects, 82 %, predicted that the predominant colour would again be correct on trial 4 (combined results for sequences 1 and 2 of condition *D*); whereas following the run of position on trials 1–3, only sixteen out of eighty, 20 %, predicted that the run would continue on trial 4 (condition *C*). The difference is highly significant ($P < 0{\cdot}001$). According to this criterion, we would conclude that runs of colour induce positive recency, whereas runs of position cause negative recency.

The response to the long run of position (trials 10–19) shows that negative recency in the present situation represents only a passing phase. Presumably subjects only predict that the run will end for so long as they maintain the hypothesis that outcomes are selected randomly with respect to position. A very long run constitutes contradictory evidence, and at some point subjects might reasonably be expected to reject this hypothesis in favour of the hypothesis that alternatives from one position are being selected systematically.

These results confirm that both positive and negative recency can be obtained in a situation which demands a simple choice between alternatives. The nature of the influences determining recency must however be even more subtle than has been supposed, since the dimension (colour or position) would appear to determine which response will occur.

Nicks (1959) has reported that negative recency occurred to runs on one of two colours when subjects had to guess whether a red or green lamp would be lit on each trial. The contradiction between the present results and those of Nicks is only apparent, however, since Nicks placed his lamps side by side, thus confounding differences in colour with differences in position. Nonetheless it must be noted that Jarvik (1951) obtained negative recency when subjects had to guess whether he would say 'plus' or 'check' on each trial, so it cannot be argued that the phenomenon is a special response to position differences.

With sequence 1 in the present experiment, position and colour were confounded on the first three trials, so that the outcomes were consistent with both rules. If subjects chose on trial 4 from the side which had previously been correct, this could indicate positive recency induced by the run of one postion or negative recency induced by the run of one colour. With sequence 2, by contrast, this early confounding was absent. A comparison of the responses to the different sequences on these early trials is therefore instructive. Considering only the results from condition *D* (targets always the same colour): with sequence 1, on trial 3, sixteen out of forty subjects or 40 % predicted correctly; but with sequence 2, thirty-one out of forty subjects or 77 % were correct on this trial (cf. Table 2). The difference is highly significant statistically ($\chi^2 = 23$, D.F. $= 1$; $P < 0{\cdot}001$). Thus with sequence 1 performance was depressed by the negative recency induced by regularity of position. The level was not as low, however, as would be expected had position been the only rule consistent with the outcomes.

An analysis of the subjects' comments is given in Table 3. On the whole, the subjects who were able to describe the rule correctly predicted correctly over the latter half of the run, but a few had predicted so poorly that they must have discovered the rule only after the run had ended. Those categorized as 'guessers' had chance level scores, their overall mean number of correct predictions being 10·2. When the rule was 'all one colour', sixteen out of the total of thirty-two who were questioned apparently failed to perceive it. When it was 'all one column', twelve of the fifty-eight who gave comments appeared not to have noticed it. One of these found red/black differentiation so compelling that he commented, 'I made it work out to more or less ½ red and ½ black numbers'. Since these comments were elicited at the end of the experimental session they refer only to the subject's ultimate response pattern. They throw no light on the way their hypothesis might have changed during the run. In the records of some subjects there was evidence that their strategy had altered during the run. Thus a number of them obviously guessed at first, and later discovered the correct rule (these subjects have been included in those knowing the rule). The supplementary experiment that is reported next was carried out to throw further light upon subjects' initial expectations in this situation.

Table 3. *Analysis of comments in Expt. II*

Rule	'All one colour'	'All one side'
Reported rule	15	20
Guessed	10	9
Sought Sequence	3	4
Others	0	8
No comment	4	17
Number solicited	32	58

EXPERIMENT III

Assessing initial expectations

The interpretation given to the reaction to a run of one position in Expt. II was in terms of the subjects' hypotheses and the way these were influenced by evidence. It is possible that the differential reaction to runs of colour and position can also be related to these hypotheses. To obtain evidence about subjects' expectations, the experiment was repeated using new subjects. This time, instead of waiting until the end to question the subjects, the session was interrupted for interrogation immediately after the first choice had been made. Two procedures were employed, first a forced choice questionnaire and secondly an open invitation to suggest hypotheses.

(*i*) *Forced-choice method*

The material was the same as in Expt. II. There were two groups: in one, half the subjects had list 1A and half list 2B; in the other, half had list 1B and half list 2A. Instructions and the initial procedure were exactly the same as for Expt. II, but after the first outcome had been announced the run was abruptly terminated. For the first group the correct target had been 43, which was on the right. For the second it had been 34, which was on the left. In both cases it was red.

Instead of continuing the run, the subjects were asked to complete the following questionnaire:

There are many ways in which I might pick the numbers on my list. The four possibilities I want you to consider are:

A I always pick the smaller numbers
B I always pick the red one
C I just choose at random
D I always pick the right-hand one (for the second group this read 'left-hand')

Rank these possibilities according to their likelihood:

The most likely is ....
The second-most likely is ....
The third most likely is.....................

A preliminary group of twenty-one subjects who were given the questionnaire exactly as printed here ranked them: random choice, constant colour, constant size, constant position, in descending order of likelihood. This order could have been influenced by response bias. The questionnaire was therefore given to another forty-two subjects, this time with the statements ordered: A, I just choose at random: B, I always pick the red one; C, I always pick the bigger (smaller) number; D, I always pick the right-hand (left-hand) one. In order to confirm the preliminary findings the preferred order would need to be A, B, C, D. It was considered that response bias would make this ranking unlikely, so that in order to confirm the preliminary findings bias would need to be overcome.

The data from the preliminary test and the re-ordered questionnaire were extremely similar and have therefore been pooled. The combined data (see Table 4) indicate that there was a fair degree of concordance between the different rankings ($W = 0{\cdot}32$; $P < 0{\cdot}01$ for both χ^2 and F tests). The differences in the mean rank order reveal that the subjects generally considered random choice to be likely while the selection of all numbers from the same side was most unlikely. The selection of items of one colour was given an intermediate rating. Fifty-one of the sixty-three subjects ranked constant colour as more likely than constant position; this is significantly more than would be expected if no real preference existed ($P < 0{\cdot}01$, binomial test). Forty-four of the sixty-three subjects ranked random choice as more likely than constant colour, which again is significantly greater than the chance level ($P < 0{\cdot}01$, binomial test).

Table 4. *Ranked probability of four possible rules after the first (and only) trial* ($n = 63$)

	Rank order				
Rule	1	2	3	4	Mean
Random choice	39	9	9	6	1·71
Constant colour	15	20	18	10	2·38
Constant size	9	17	27	10	2·60
Constant position	0	7	19	37	3·48

(*ii*) *Open questioning*

The general method was the same as with the forced-choice questionnaire. The guessing task was run for one trial, but then instead of being given the questionnaire the subjects were asked to describe in writing the way in which they thought the experimenter was selecting the target numbers. After 3 min. they were then invited to report a second hypothesis and a third, labelling the first as 'most likely', the next as 'second most likely' and so on.

A total of thirty-one subjects participated. With the exception of a few rather odd hypotheses, the rules they suggested were of limited variety and could be tabulated (see Table 5). As with the forced-choice method, constant colour is considered to be much more likely than constant position. The only difference of any magnitude between the results of the two methods is that with the open question random choice is rarely mentioned, whereas it ranked most highly in the questionnaire. This can possibly be attributed to the set imposed by the instructions. The suggestion that the experimenter could choose at random had been deliberately avoided to prevent biasing the outcome and it is possible that the instructions unwittingly implied that some rule was being employed.

Table 5. *Rules suggested by subjects after their first* (*and only*) *trial, showing the number of times each rule was suggested and the order in which it was given* ($n = 31$)

Rule	Order of 'probability' 1	2	3	Total (including lower rankings)
Constant colour	17	4	1	23
Alternating colour	2	2	0	5
Constant size	5	6	2	13
Alternating size	0	1	1	2
Constant position	1	2	2	6
Pattern or alternating position	2	8	3	16
Random	0	1	1	3
Other	4	2	4	10
Total	31	26	14	78

The odd hypotheses described by some subjects are of interest as curiosities. One subject thought that difficulty of enunciation could be the basis of the rule. Another thought that some numbers were more 'eye catching' than others and that these would be chosen. A third believed that the experimenter's personal choice or else that of his superiors was used.

GENERAL DISCUSSION

This series of experiments has shown that the factors which determine the reaction to regularity in two-choice guessing differ from those previously thought to be important (Dale, 1965). From the results of the earlier experiments it had appeared that the regularity of the clue-target relation, as used in Expt. I, was the crucial factor in inducing positive recency. But it is not. Positive recency can be induced using a simple choice between pairs of stimuli without the complication of clues.

With the multidimensional stimuli of Expt. II, subjects were more willing to accept the hypothesis that the correct alternative was always the same colour than they were to accept a constant position hypothesis. In this experiment runs of both position and colour were easy to detect; it is therefore unlikely that the difference in response can be attributed to ease of detection. Instead, the difference appears to depend upon the relative probability, in the subjects' own estimation, of the alternative hypotheses. The hypothesis that the correct numbers would all be the same colour was judged to be relatively likely. According to Bayesian decision processes (Edwards, Lindman & Savage, 1963) a rational person requires relatively little evidence to be convinced of the truth of a hypothesis with a high *a priori* probability. The readiness with which subjects reacted to the runs of one colour as evidence that an all-one-colour rule was being employed is therefore reasonable. No attempt at precise measurements of the subjective probability of the alternative hypotheses was made, so this indication that subjects behave in accordance with Bayesian decision rules is only qualitative. The rapidity with which positive recency develops as a response to a run of one colour is striking, however, and it contrasts with the conservatism in using evidence which has been shown in studies directed at assessing the efficiency of subjects as Bayesian operators (Edwards & Phillips, 1964).

Although regular selection from one position was generally regarded as unlikely, subjects appeared to accept the rule when sufficient evidence had accumulated. Their reaction to a long position run thus included a prolonged phase of negative recency, followed eventually by positive recency. Typically there was an additional short-lived period of positive recency at the beginning of the run so that the complete reaction showed three phases. Signs that the reaction to a long position run is complex and of this form also exist in other reports of two-choice guessing experiments, notably in those of Nicks (1959) and Lindman & Edwards (1961).

The experiments reported have shown that markedly different reactions are made to different forms of regularity. A similar finding has been reported by Anderson (1960) who noticed positive recency to regular alternation of position, whereas negative recency was obtained with constant selection from the same position. The questionnaires have revealed that the different reactions are related in a sensible way to the subjects' expectations. Subjects think it likely that the experimenter will select all the red numbers, but most unlikely that he will always select from the same position. Although it is comforting that such a rational relation between behaviour and stated expectations should exist, this correlation does not constitute an explanation of the subjects' behaviour since it leaves unanswered the question of why the subjects should expect systematic choice of one colour, but not of one position.

REFERENCES

ANDERSON, N. S. (1960). Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.* **59**, 73–93.

DALE, H. C. A. (1965). Recency in two-choice guessing tasks. *Brit. J. Psychol.* **56**, 33–44.

EDWARDS, W., LINDMAN, H. & SAVAGE, L. J. (1963). Bayesian statistical inference for psychological research. *Psychol. Rev.* **70**, 193–242.

EDWARDS, W. & PHILLIPS, L. D. (1964). Man as a transducer for probabilities in Bayesian command and control systems. In *Human Judgements and Optimality*, ed. M. W. Shelley and G. L. Bryan. New York: Wiley.

GREEN, R. T. (1958). Factors affecting inductive predictions. *Brit. J. Psychol.* **49**, 97–105.

JARVIK, M. E. (1951). Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.* **41**, 291–7.

LINDMAN, H. & EDWARDS, W. (1961). Unlearning the gambler's fallacy. *J. exp. Pyschol.* **62**, 630.

NICKS, D. C. (1959). Prediction of sequential two-choice decisions from event runs. *J. exp. Psychol.* **57**, 105–14.

(*Manuscript received* 22 *September* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 45–52
Printed in Great Britain

VARYING THE NUMBER OF ALTERNATIVES IN SHORT-TERM RECALL

BY MURIEL M. WOODHEAD
Medical Research Council, Applied Psychology Research Unit, Cambridge

Performance on a task of continuing memory span improved as the number of alternative items increased, in contrast to short-term recall from single presentations. The tentative explanation offered is that the difficulty of organizing items in chunks during continuing recall increases as the amount of information decreases.

To begin with a conclusion, we deduce from many studies of memory span that it is more difficult to recall items selected from large, rather than small, ranges of alternatives. Developing this position Miller (1951) suggested that the amount of information stored, not the total number of items, limits the span of immediate memory. The amount of information conveyed by an item is a function of the number of its possible alternatives. But later Miller (1956) mentioned that if items can be grouped in chunks, the information within a chunk is less important than the number of chunks. Therefore a reduction in information can only be helpful if it allows the material to be memorized, to be so organized, and fortunately this does apply to most of the things we wish to remember as, for example, the letter structure of words.

Among the conclusions drawn by Pollack (1953) was that, for messages of a given length, the information lost is independent of the number of possible alternatives. Subsequently, in an experiment on continuing memory span, Pollack & Johnson (1963) required random digits to be recalled sequentially after a delay for each item, equal to a fixed number of digits, in conditions which included two, three and four alternatives; and were able to conclude that the estimate of short-term informational storage does not depend on the number of alternatives.

Conrad & Hille (1957), giving a series of tests of the immediate memory span, reported that telephone operators reading digits reproduced decreasing vocabularies with increasing accuracy. Recently Conrad (1964) described a memory experiment using letters of the alphabet. (Twenty-six letters must, of course, have higher possibilities of confusion than 10 digits.) Errors selected for analysis were single occurrences of a wrong letter within a sequence; all others were excluded. From the results Conrad predicted that memory span is a function of the number of acoustic confusions expected from the vocabulary of items used.

The present study attempted to investigate continuing memory span, with varying amounts of alternatives. The results are explicable in terms of Miller's chunk theory, but since the number of alternatives proved a significant factor, they differ from those of Pollack & Johnson (1963). Less definitely, there is also some slight negative evidence in the matter of acoustic similarities.

The main experiment, on continuing memory, was preceded by a short-term single-presentation memory test having two and six stimulus alternatives; its purpose was to demonstrate that, for the subsequent use of this test material, the smaller range of alternatives was the easier; a secondary interest being whether the type of error was a factor in any difference due to the number of alternatives. For the main experiment, a continuing memory test was devised with two, six and ten stimulus alternatives. Here it was hoped to learn whether the accepted order of increased difficulty with increased alternatives holds for a type of memorizing in which it could be difficult to organize items in chunks.

SINGLE-PRESENTATION MEMORY: EXPERIMENT I

Method

Subjects. This test was given to eighteen sailors, aged between 17 and 27 yr.

Construction. On each of 180 cards six digits were typed centrally. Ninety of the series of six were all different, but always included the numbers 2, 3, 4, 5, 8, 9, each occurring with the same frequency in every serial position; this was the 6-alternative condition. The remaining ninety cards contained two alternatives, the numbers 2 and 3. Neither of these numbers appeared more than three times in succession within any series of six. Since with this restraint it was possible to construct forty-eight series, forty-two were presented twice, not on the same day or in the same order. The frequency for each number in every serial position varied between forty-four and forty-six.

Easily discriminated numbers had been avoided, but it is not claimed that those chosen form the most confusing combination possible. In a table prepared by Moser & Fotheringham (1960) people from five different native-language backgrounds heard and responded to 21,000 digits through noise and the highest confusion is listed as response '3' to stimulus '2'. For the present investigation eighteen native English speakers (not taking part in the two experiments) made 1900 responses to each digit, also mistaking 2 for 3 more than any other pair. When their errors were combined to form all possible sets of six digits from ten, the highest frequency occurred for 2, 3, 4, 5, 8, 9. Although it does not accord with the highest obtainable from the matrix of Moser and Fotheringham, this set was preferred because it came from the language population to which it would be applied.

Procedure. Subjects were individually tested, and given both conditions of alternatives. A card was exposed for 2 sec. The man then wrote the six digits at his own speed. Successive presentations of thirty cards were randomly and differently distributed for every subject, but always contained fifteen of each alternative. There was an interval of about 3 min after the 30th, 60th, 120th and 150th responses. The test was given in two parts, each of ninety cards, on successive afternoons.

Results and discussion

Significantly more correct digits were recalled with two alternatives than with six ($P < 0{\cdot}01$), but the difference disappeared when the scores were corrected for random guessing. (Reasons for preferring the random guessing rule to the constant ratio rule appear at the end of the paper.) It is possible that every wrongly reproduced digit was the consequence of its acoustic similarity to another, particularly since the more confusable digits were deliberately selected for the test. If this is the explanation, the likelihood of confusing one digit with another would have seemed to be independent of the number of alternatives.

Further inspection of the errors, however, revealed three distinctions: series reproduced in full but not in the correct order (e.g. 589234 for 598234, and 332332 for 332323); substituted digits (e.g. 452358 for 452398, and 322333 for 322323); and intrusions and omissions. Instances of the last two categories were too few for analysis. Both order and substitution errors could have arisen from acoustic similarities. In the case of substitutions, the frequency with which specific digits replaced each other was associated with the frequency pattern of the eighteen native-English speakers, mentioned above (single-tail $P = 0{\cdot}038$ for difference of

Kendall's τ from zero), but not with the frequencies of Moser & Fotheringham (1960). Order errors were not related to either pattern. Therefore it was considered there existed some difference other than, or as well as, any acoustic similarity. The results of analysing the two main types of error appear in Table 1. Table 2 presents the group means.

Table 1. *Single-presentation memory experiment: differences between two and six alternatives*

	Before correction		Corrected for guessing*	
	z	P†	z	P†
Order errors	2·93	< 0·01	2·77	< 0·01
Substitutions	1·24	N.S.	0·38	N.S.
Total wrong digits	2·63	< 0·01	0·72	N.S.

* Adjusting six alternatives to proportionate errors for two alternatives by $nq/2(n-1)$, where n is the number of alternatives and q is the observed proportion of errors.

† Wilcoxon signed-ranks statistical test.

Table 2. *Single-presentation memory experiment: mean errors*

	Alternatives	Before correction	Corrected for guessing
Order errors	2	6·06	6·06
	6	12·89	7·73
Substitutions	2	11·61	11·61
	6	21·83	13·48
Total wrong digits (max. 1080)	2	27·00	27·00
	6	50·00	30·00

The direction of the differences was towards superior performance with the smaller range of alternatives, in agreement with studies summarized by Miller (1951) and the finding of Conrad & Hille (1957). The error breakdown demonstrates this was due to order errors, and this held regardless of the number of 2's and 3's within each series in the 2-alternative condition. This order error finding is in contrast to Conrad's (1964) prediction of acoustic confusion. But Conrad also stated that confusion refers to recall whether acoustic or not, and since his prediction followed a long series of memory studies and their place in the decay theory of immediate memory, it is assumed that the present order errors are confusions during retrieval from storage, the digits having been perceived in their correct order. In a similar context, Moray & Barnett (1965) concluded that order errors in short-term memory are a measure of performance with good perception and poor recall.

It is certainly not unreasonable to have a class of order errors, nor even to find them dominant. Broadbent (1965) comments that if the order in which items occur is stored separately from the items, then order information must be treated separately from item information. Brown (1958) assumes the high probability of forgetting order-information, suggesting that it fades more rapidly than identity-information. The effect of order errors, he states, is far-reaching because items recalled in the wrong order automatically shorten and lengthen the delay before recalling the other items in the set; those delayed unduly will be more liable to be forgotten. For short-term memory, this may mean that in an error hierarchy, order errors lead. Although it seems scarcely possible to arrive at the cause of the confusion which results in order errors, it may be distinguishable from the confusion causing other errors, by its location, either in the act of silent rehearsal (with the subject transposing items as he repeats them to himself), or in the act of responding (with the subject rehearsing correctly but transposing the items as he writes or speaks). Not even slight circumstantial evidence of this kind could be available in the case of substitution errors. It is theoretically possible that confidence ratings for each digit would yield smaller degrees of confidence for substitutions than for order errors, because the latter are constituted of correct digits.

However, the immediate value of this experiment was the indication that digit vocabularies of this kind would be suitable for demonstrating at least some alterations in response with varying ranges of stimulus alternatives under the condition of continuing memory span.

CONTINUING MEMORY SPAN: EXPERIMENT II

Method

Subjects. Forty-eight housewives aged between 25 and 50 yr performed the test.

Construction. Thirty-two sequences of sixty items were typed vertically on paper strips. Sequences 1–8 were composed of the numbers 2 and 3, sequences 9–16 of the numbers 2 and 4 (a contrasting pair which had produced a low confusion reading), sequences 17–24 of the numbers 2, 3, 4, 5, 8, 9, and sequences 25–32 of the numbers 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. These supplied three ranges of two, six and ten alternatives. Digits occurred equally often within a sequence. Adjacent like numbers (repetitions) occurred equally often in all sequences within each alternative, but not between alternatives; at that time studies of the role of repetitions by Conrad (1965) and Wickelgren (1965) had not been published; in fact sequences with two alternatives contained five doubles and two triples, six alternatives contained five doubles, ten alternatives contained three doubles.

The presentation apparatus consisted of a belt, stepped by a solenoid controlled by an electronic timer. The paper strip fitted over the belt which moved under a cover having an aperture to reveal each digit singly together with space for a written response.

Procedure. The paced task required a subject to view successive digits and recall the first one when the fourth appeared, the second when the fifth appeared, and so on through a sequence. Subjects were advised to guess rather than leave a gap. Different subjects were used in each condition, mainly because the task required considerable learning. The subjects were told the actual two, six or ten digits which would appear. On the first of two successive afternoons, an individual subject had four training sessions, as follows: (1) 1-back recall, with intervals of 2 sec for combined exposure of the current item and response to a previous item, (2) 2-back recall at 3 sec intervals, (3) 3-back-recall at 4 sec intervals, (4) 3-back recall at 3 sec intervals. Four sessions 24 hr later consisted of 3-back recall at 3 sec intervals. The first five of the eight sessions were discarded, the last three analysed as experimental data. Since it was not until the seventh digit that subjects were remembering three items in sequence which had been perceived while already remembering three items, scoring for 3-back was based on the fifty digits which were the seventh through the fifty-sixth.

Results

Two of the four possible types of error could be identified with certainty: omissions; and out-of-place errors, one of the types of order error, when subjects were recording 1-, 2-, or 4-back instead of 3-back, for instance. The remaining errors were apparently substitutions and order errors, the latter in 2-alternative conditions only; but in many instances it was difficult to be sure of their category. All types of errors were therefore combined, and the group means appear in Table 3.

In Table 4 can be seen the occurrence of some differences between conditions of alternative. Their direction, however, was not in the accepted order of difficulty with rising number of alternatives. Under these circumstances a random guessing correction enlarged the region of rejection of no effect.

Table 3. *Continuing memory span experiment: mean errors* (*max.* 150)

Alternatives	Before correction	Corrected for guessing
2 (High confusion)	52·25	52·25
2 (Low confusion)	56·6	56·6
6	34·75	20·85
10	27·83	15·45

Table 4. *Continuing memory span experiment: differences between numbers of alternatives for high confusion* (*HC*) *and low confusion* (*LC*) *series*

	Before correction			Corrected for guessing		
Pairs of alternatives	*z*	*P**	Direction of errors	*z*	*P**	Direction of errors
2(HC) × 2(LC)	0·03	N.S.	2(HC) = 2(LC)	0·03	N.S.	2(HC) = 2(LC)
2(HC) × 6	1·71	< 0·05 (1-tail)	2(HC) ⩾ 6	2·49	< 0·02	2(HC) > 6
2(LC) × 6	1·53	N.S.	2(LC) = 6	2·37	< 0·02	2(LC) > 6
2(HC) × 10	2·14	< 0·05	2(HC) > 10	2·77	< 0·01	2(HC) > 10
2(LC) × 10	2·11	< 0·05	2(LC) > 10	2·92	< 0·01	2(LC) > 10
6 × 10	0·09	N.S.	6 = 10	0·14	N.S.	6 = 10

* Mann–Whitney *U* test.

Further analysis, summarized in Table 5, indicated that differences occurred in ranked order of alternatives. The more alternatives there were, the more performance improved (Fig. 1).

Table 5. *Continuing memory span experiment: the ordered effect of three alternatives for high confusion* (*HC*) *and low confusion* (*LC*) *series*

	Before correction		Corrected for guessing	
Groups	*z*	*P**	*z*	*P**
2(HC), 6, 10	1·89	< 0·05	2·67	< 0·01
2(LC), 6, 10	1·78	< 0·05	2·68	< 0·01
2(HC), 2(LC), 6, 10	2·38	< 0·01	3·39	< 0·001

* Jonckheere's test against ordered alternatives.

Discussion

Measurement of the order of differences between alternatives confirmed that efficiency was rising as alternatives increased. Direct comparison of the two larger ranges, however, had failed to reveal a difference between them. Perhaps this was due to the experimental framework, for instance the prolonged practice; but a more probable explanation may lie in the comments of the 6- and 10-alternative subjects, all of whom reported using various 'systems', whereas subjects working with two alternatives inquired if there existed any method to help them do the test; the experimenter always replied: 'I'm not sure. Do tell me if you find one.' All subjects indicated conversationally their search for, or adoption of, a working routine. Whether this is interpreted as a need to direct most of their immediate attention to identifying the digits or not, it must be concluded that subjects wished and attempted to organize their working material.

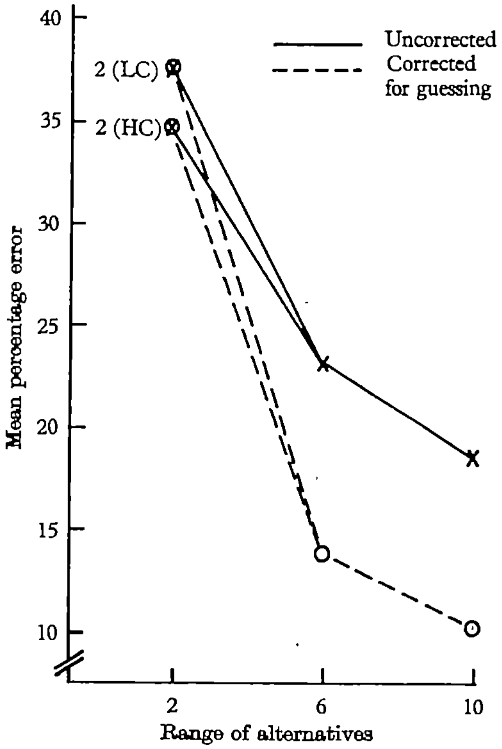


Fig. 1. Continuing memory span: effect of increasing the range of alternatives.

If, despite the special memory requirements of continuing recall, subjects were able to organize the wider ranges of alternative digits in chunks, their lack of success with two alternatives was not evidently of an acoustically confusing nature, otherwise there should have appeared at least a directional tendency towards a difference between the two 2-alternative conditions, in favour of the less confusing one. That this was not the case, may be observed from Fig. 1.

On the other hand, there must be a doubt as to how much the greater frequency of repetitions with two alternatives altered the level of difficulty for organizing successive digits in groups. It will be recalled that in the earlier single-presentation memory experiment, either lack of repetitions did not facilitate the 6-alternative condition, which proved harder than two alternatives, or the latter condition was made easier by the unavoidable repetitions. Taken together, the most recent studies of repetitions give little guidance. The differing structures of repetition information in two sizes of letter vocabulary had no effect on performance (Conrad & Hull, 1964). But when five lengths of digit sequences were recalled, memory for the *non*-repeated items in sequences containing some repetitions was better than for corresponding items of all different sequences (Wickelgren, 1965). Again, given four repetition-types of digits sequences, the best performance came from sequences without repeats, and recall deteriorated as repetitions increased (Conrad 1965). Since the findings of these memory studies run counter to one another, and did not in any case include continuing memory span, it is evident that here is a gap to be explored. For the present it may be pointed out that repetitions might involve the slight extra task of remembering how many times an item was repeated. Consequently, if repetitions facilitate recall, the extra task detracts from the facilitation, or if repetitions confuse the subject the confusion could consist of remembering their number or it may be a further addition to order and acoustic confusion and increase the overall difficulty.

In comparison with other aspects of the task, repetitions may be a minor factor. Attempting to organize the material into cohesive chunks during a continuing memory span operation is a hazardous venture when one considers that immediately following a response the item concerned must be discarded from the chunk and a new item added in a different serial position, while at the same time every other item has shifted its serial position, inhibitory procedures which involve almost simultaneous disintegration and assembly of chunks. Obviously a single chunk can never be transferred as a whole from one response time unit to the next. It can hardly be surprising if continuing memory span is not independent of the amount of available information. With straightforward memorizing, the economical system of chunking is undisturbed, so that when alternatives arise, it is easier to discriminate among few rather than many; but where there is a constant manœuvring of the items which should be formed into chunks, people perform better with more alternatives. Thus they apparently depend on having more information, possibly as an aid to discrimination. Though it cannot be more than a tentative conclusion, it does seem so far that reducing the amount of information relative to alternatives hampers the formation of chunks and thus limits the continuing memory span.

The direct observation from this study, however, was that continuing memory span for digits functioned conversely to single-presentation memory span, the smaller ranges of alternatives proving more difficult than the large.

DISCUSSION

Corrections for guessing

In both experiments, data from the larger ranges of alternatives were adjusted for chance effects to compare with the smallest range. But we have subjects' assurance that they pursue a system, thus reducing the likelihood of guessing, particularly in the larger ranges. The random guessing rule (RGR) assumes the subject remembers correctly or guesses, whereas he may actually remember correctly, guess, *and* respond incorrectly but systematically. The RGR caters for some of the errors but is applied to the whole. However, putting this restraint on the data obtained from the single-presentation memory experiment did not alter its tendency to fit established patterns of recall.

The weakness of the RGR prompted an alternative correction by the constant ratio rule (CRR), which assumes the ratio of response frequencies is independent of the number of alternative responses available. Application of the CRR to the scores of the short-term memory experiment (designed not to uncover new aspects of behaviour but to confirm the suitability of test material) reversed the trend of performance, producing significant results inconsistent with the well-attested effects of increasing alternatives. The results of the RGR adjustment were more plausible. The problem scarcely existed for the continuing memory span experiment, because both kinds of correction increased the effects owing to the direction of the errors, 2 > 6 > 10, which is opposite to the direction which a random answering machine would automatically produce.

The author wishes to thank Dr Donald Broadbent who directed the study and Mr Peter Freeman for his statistical advice.

REFERENCES

BROADBENT, D. E. (1965). Techniques in the study of short term memory. *Acta Psychol.* **24**, 220–33.

BROWN, J. (1958). *Information, Redundancy and Decay of the Memory Trace.* Teddington, Middlesex: National Physical Laboratory.

CONRAD, R. (1964). Acoustic confusions in immediate memory. *Brit. J. Psychol.* **55**, 75–84.

CONRAD, R. (1965). The role of the nature of the material in verbal learning. *Acta Psychol.* **24**, 244–52.

CONRAD, R. & HILLE, B. A. (1957). Memory for long telephone numbers. *P.O. Telecom. J.* **10**, 37–9.

CONRAD, R. & HULL, A. J. (1964). Information, acoustic confusion and memory span. *Brit. J. Psychol.* **55**, 429–32.

MILLER, G. A. (1951). *Language and Communication.* New York: McGraw-Hill.

MILLER, G. A. (1956). The magical number seven—plus or minus two. *Psychol. Rev.* **63**, 81–97.

MORAY, N. & BARNETT, T. (1965). Stimulus presentation and methods of scoring in short-term memory experiments. *Acta Psychol.* **24**, 253–63.

MOSER, H. M. & FOTHERINGHAM, W. C. (1960). Number telling. *U.S.A.R.D.C. Tech. Rep.*, no. 58, *RF Project* 1080.

POLLACK, I. (1953). Assimilation of sequentially encoded information. *Amer. J. Psychol.* **66**, 421–35.

POLLACK, I. & JOHNSON, L. (1963). Continuing memory span for digits. *Percept. mot. Skills*, **17**, 731–4.

WICKELGREN, W. A. (1965). Short-term memory for repeated and non-repeated items. *Quart. J. exp. Psychol.* **17**, 14–25.

(*Manuscript received* 26 *October* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 53–59
Printed in Great Britain

WORKING BACKWARD AND WORKING FORWARD IN PROBLEM-SOLVING

By W. S. ANTHONY
Institute of Education, University of Hull

N.E. Miller (1959) and others have stressed the effectiveness and creativity of working backward in problem-solving; but Newell, Shaw, & Simon (1962) argued that working backward is not inherently superior to working forward. An experiment is reported in which undergraduate subjects were set a series of route-finding tasks in a maze-situation modified to allow observation of working forward and working backward. The effectiveness of the two directions of work was controlled. The method and results of the experiment are discussed in relation to the views of the above-named writers and others.

'Working backward' is a method of problem-solving usually contrasted, explicitly or implicitly, with 'working forward'. For the purpose of applying the concepts, a problem is conceived as the task of finding a method of proceeding from a given situation or starting-place to a required situation or goal. To work forward then means, roughly, to work from the starting-place, extending one's knowledge of routes leading out from the starting-place; similarly, to work backward means to work from the goal, extending one's knowledge of routes leading into the goal. The concepts of working forward and working backward are attractive because they can be applied to many and perhaps to all problem-situations; they might be useful subordinate classes of the general concept of 'direction' in problem-solving (Maier, 1930; Bartlett, 1958; Berlyne, 1965).

Dollard & Miller (1950), Miller (1959) and Polya (1945) presented anecdotal or fictional examples which show working backward to be relatively effective and intelligent in contrast to working forward. In Miller's exposition, working backward is a type of 'creative thinking'; working forward may be 'creative', if guided by the goal, or may be 'mere symbolic trial and error'. Aristotle (1947, p. 52; cf. Heath, 1949), Pappus (cf. Heath, 1908 and Polya, 1945), and Berlyne (1965, pp. 301–2) have also written in apparent commendation of working backward.

Newell *et al.* (1962) argued that working backward is more effective than working forward in problems of proof-finding in logic owing to the structure of such problems, but they stressed that working backward is not 'inherently superior' to working forward. Van de Geer (1957, pp. 161–2) presented a similar analysis. Newell *et al.* (1962) were particularly concerned with the simulation of human problem-solving by computer programmes; they reported a few records of human subjects 'thinking aloud' while attacking proof-finding problems.

Duncker (1945) reported several experiments in which subjects 'think aloud' while attacking problems. Duncker distinguished various methods of attack, among which 'analysis of the situation' (which includes working forward) and 'analysis of the goal' (which includes working backward) were prominent; on the whole, he did not appear to value one of these methods more than the other, although in one passage, mentioned by Newell *et al.* (1962), working backward was given a status superior to working forward (Duncker, 1945, p. 44). Merrifield, Guilford, Christensen & Frick (1960) report a factor-analytic study of problem-solving abilities. One of the authors' conclusions was that the kind of test that seemed most related to the criterion problem-solving tests was the kind of test designed to measure abilities to work from goal information.

The present paper reports an experimental study of working forward and working backward. The experiment was undertaken with three main intentions: (i) to obtain simple observations of working forward and working backward, (ii) to illustrate an implication drawn from Van de Geer (1957) and Newell *et al.* (1962), that the relative effectiveness of the two directions of work may be determined by the experimenter, and (iii) to test for adaptive tendencies in the subject's direction of work. A maze or route-finding situation was used, because it is convenient and because it has been regarded by such writers as Miller (1959), Newell *et al.* (1962), Berlyne (1965, p. 299) and Woodworth & Schlosberg (1955) as a basic analogy for, or typical example of, problem-solving. Because normal maze situations, including the sectional map-reading task of Bartlett (1958), are designed for the observation of working forward only, a modified task of the following kind was constructed.

The subject is given a paper on which two circles, labelled 'city A' and 'city B', for instance, have been drawn. This paper is an incomplete map of an imaginary country. The subject's task is to obtain a map of the route from city A to city B. There is only one route. He begins by writing an 'X' in one of the circles; then the experimenter completes the part of the map in the vicinity of that circle and the subject writes an X in another circle, perhaps in a circle recently added by the experimenter; the map is completed in the vicinity of that circle, and so on until the map is sufficiently complete to show the route from city A to city B. The subject is given six tasks of this form.

The application of the concepts of working forward and working backward to this situation is justified by the instruction to obtain a map of the route from one place, the 'starting-place' of the problem, to another place, the 'goal' of the problem. Every response that the subject makes can be classified unambiguously as a working forward response or a working backward response.

METHOD

The tasks

Two types of task were used, both of the general form described above. In tasks of type 'F', working forward was relatively effective, and in tasks of type 'B', working backward was relatively effective. In a task of type F the country is of the form shown in Fig. 1A, but, at the beginning of the task, only the two circles representing the start and the goal, and their labels, 'city A' and 'city B', appear on the paper. If the subject consistently works backward from the goal to the start he will tend to waste responses in exploring blind alleys, but if he consistently works forward from the start to the goal he will not waste responses. On the other hand, a subject who is given a task of type B (cf. Fig. 1B) will tend to waste responses if he consistently works forward; but if he consistently works backward he will not waste responses. For these reasons, a working forward response is here termed 'directionally correct' for tasks of type F, and a working backward response is termed 'directionally correct' for tasks of type B. The situation was so arranged that if a subject made two or three directionally correct responses in his first three responses to any task, he completed the test by those three responses; but if he made only one or no directionally correct responses in his first three responses he had to make more responses to complete the task.

The start of the route was always represented by the lower of the two circles initially drawn; the end of the route, by the upper circle. As the paper lay on the table in front of the subject, the 'lower' circle was nearer to him; 'working forward' was therefore forward in space as well as 'from the start', and 'working backward' was backward in space as well as being 'from the goal'. In each task the route consisted of three sections of road; of the four cities in a straight row, only one was on the route. The middle section of the route varied from task to task, within each type of task; it always ran from the one city which lay in a straight row with the start and goal, to some one of the cities in the straight row of four.

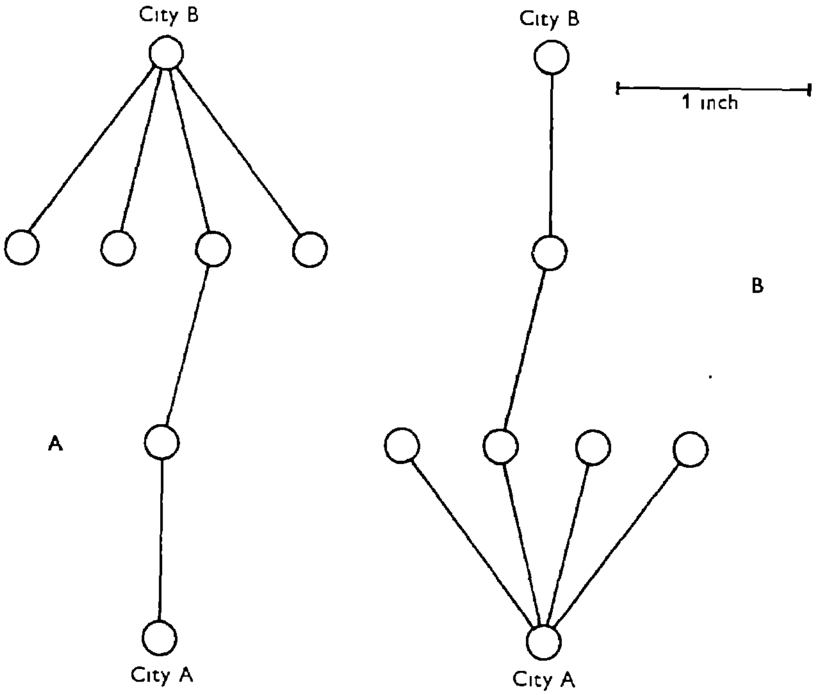


Fig. 1. A, the type of country to be discovered in tasks of type F; B, the type of country to be discovered in tasks of type B. The scale line and words '1 inch' are for the reader of this article, to show the size of the subjects' maps, and did not appear on the maps themselves. The words 'city A', 'city B' (or C and D, etc.) did appear on the maps, as explained in the text. The circles and lines were traced on to the maps freehand from a master-map. The middle section of the route varied, as explained in the text.

Plan of the experiment

Two groups of subjects were used. Subjects were set five tasks of one type and then a sixth task of the other type. The first five tasks are called the 'training tasks', and the sixth task the 'reversal task'. Subjects in group 1 were given five training tasks of type F and a reversal task of type B; subjects in group 2 were given training tasks of type B and a reversal task of type F.

Subjects

Twenty first-year undergraduate students of psychology were divided into the two groups, 1 and 2. Each group contained eight men and two women. Subjects were allotted to the groups alternately within sex, with one reversed alternation to achieve a match in respect of time of day at which the subjects were tested.

Procedure

Each subject was tested individually, sitting at a table. A sheet of quarto white paper, placed in front of him, showed two circles labelled 'city A' and 'city B'.

The instructions, played from a tape recorder, told the subject that the paper was an incomplete map of an imaginary country, and that his task would be to obtain a map showing the route from city A to city B by road. The instructions went on: 'Now I must explain how you are to obtain a map of the route from city A to city B. I shall give you a pencil, and you will write an X, a letter X, inside one of the circles. This X means, "What are the roads here?" Then you will pass the paper to me, and I shall draw on it a line to show each road at that city; each line will only go as far as the next city along the road, and I shall draw a circle to show this next city. Then I shall pass the paper back to you, and you will again write X in any one of the circles on the paper, and then I shall draw the roads at that city; and we shall continue in this manner, until you have a map showing the route from city A to city B, and then your task is finished.' The instructions continued by telling the subject that all the roads were open in both directions of travel, and that there was only one route from city A to city B. He was reminded that he was not trying to map all the roads and cities in the area, but only the route from city A to city B; and he was told that there would be a series of similar tasks to which the same instructions would apply. After a second reading of the instructions the subject was given a typewritten copy and a pencil, and was asked to begin.

During each task the faces and writing movements of subject and experimenter were hidden from each other by screens. When the map of the route A–B was completed it was passed to the subject with a standard comment and withdrawn after about 10 sec. A second paper was then presented with two circles labelled 'city C, and 'city D'; the subject was set to obtain a map of the route from C to D in accordance with the same instructions which were not repeated. In this way six tasks in all were set to each subject. Between the third and fourth tasks, a written message told the subject how many X's he had written in completing each of the first three route-maps, and also told him: 'Please try to keep small the number of times that you write X, so that you would obtain each route-map after writing X only three times. This can always be done, but you may have to think about it.' The remaining three tasks were then given. Questions of administrative details were answered directly; questions of a more searching kind received uninformative replies.

An error was defined as writing X in a circle which *might* be off the route. If X was written in a circle which *was* off the route, the experimenter announced, 'There are no more roads at that city', as he returned the paper to the subject. In fact, the first error in any task was always a wasted response, as explained below. The consequence of an error was not left to chance, but was fixed by non-random variation of the route, in order to standardize the 'luck' of the subject (cf. Bruner, Goodnow & Austin, 1956, pp. 119–24). In the first task, so long as the subject went on making errors, three of the four circles in the row had to be inscribed with X before the middle section of the route was drawn; in the second task, two of the four circles; in the third task, four; in the fourth task, two; in the fifth task, three; in the sixth task, four. The subject was therefore rather 'unlucky' if he made errors. If, after making errors, the subject approached the middle section of the route from the more effective direction, the position of this section was randomly selected from the remaining possibilities.

RESULTS

If directionally correct and directionally incorrect responses are equally likely, a subject should make no or one directionally correct response as often as he makes two or three directionally correct responses, out of the first three responses to a task. On this null hypothesis only half of the subjects in a group should make two or three directionally correct responses, out of the first three responses to any given task.

Table 1 shows, for each task, the number of subjects who made two or three directionally correct responses in the first three responses to the task, and thus solved the task without errors and in the minimum number of moves. Using a two-tailed binomial test the null hypothesis is rejected in all cases when both groups are combined, except in respect of the first task. Hence there was a tendency for subjects to respond in the correct direction, both in the training tasks and in the reversal task for which the correct direction was changed.

The data in the table suggest a trend of increasing directional correctness from the first to the fifth task, and a groups-by-tasks interaction effect especially in respect of the fifth and sixth tasks. However, these suggestions are not supported by non-parametric methods of analysis; the results of a parametric analysis are dubious, owing to skew and heterogeneity of variance in the data.

The performance of group 1 in the training tasks was neither perfect nor significantly better than that of group 2: some subjects made errors by working backward when working forward would have shortened their tasks. Changing direction within a task was quite common. Out of the twenty subjects, eight always changed direction within a task, and a further eight sometimes did so.

On the very first response, before there was any difference in the treatment of the two groups, sixteen subjects worked forward and four worked backward; this is a significant majority in favour of working forward ($P = 0{\cdot}012$, two-tailed binomial test).

Table 1. *Number of subjects making two or three directionally correct responses in the first three responses to each task, and thus solving the task in the minimum number of moves*

Task ...	1st	2nd	3rd	4th	5th	6th (reversal)
Group 1 ($n = 10$)	7	8	8	9*	9*	6
Group 2 ($n = 10$)	7	7	7	7	8	10**
Total ($n = 20$)	14	15*	15*	16*	17**	16*

* $0{\cdot}05 > P > 0{\cdot}01$. **$0{\cdot}01 > P > 0{\cdot}001$

The binomial tests of significance are explained in the text.

DISCUSSION

In the present experiment the maze situation was modified in order to obtain simple observations of working forward and working backward, to control the effectiveness of these two directions of work and to test for adaptive tendencies. The modification is thought to be valuable in showing how the maze-situation may more nearly represent those problem situations in which working backward is a possibility. Statistically significant tendencies to work in the more effective direction were obtained.

Subjects often changed their direction of work in mid-problem. Newell *et al.* (1962) appeared to assume that working backward is more effective than working forward in logical proof-finding tasks; there should therefore be no 'need' for a 'programme' or a human subject to change direction. However, one protocol reported by them (Newell *et al.* 1962, p. 94) seems to show a clear change in direction. If such changes are common in proof-finding by human subjects as in the present results, this aspect of behaviour would have to be simulated by any programme which was presented as a theory of human problem-solving.

In some previous discussions of working backward, noted in the introduction, especially in the theory of Miller (1959, p. 245), the possibility of uncreative or ineffective working backward is ignored. It is therefore of theoretical interest that some subjects in group 1 made errors in the training tasks, since these errors consisted in ineffective and relatively stupid working backward. Thus, 'mere trial and error' may appear in *both* directions of work.

In the introduction it was suggested that, by definition, working forward involves finding routes leading out of the starting-place and that working backward involves finding routes leading into the goal. The terms are most appropriate when it is implied that a solution is required for a specified direction, i.e. from a specified starting-place to a specified goal. There may then be a tendency to work forward while finding the solution, in anticipation of its achievement. Such an effect is suggested by the tendency of subjects to work forward in making their first responses.

In the present experiment, as all roads were open in both directions, 'leading into' is equivalent to 'leading out of'; hence working forward and working backward need not be regarded as finding different *types* of route. Heath (1908, p. 139), discussing elementary Euclidean geometry, noted that if A implies B, it will often be true that B implies A; consequently, one may work backward deductively, and then verify the proof by working forward deductively. Both directions of work may be deductive. However, Newell, Shaw & Simon (1958), discussing problems where often A implies B but is not implied by B, describe working forward as identifying only propositions which are implied by the premises, working backward as identifying only propositions which imply the theorem to be proved. Such a difference between working forward and working backward was minimized in the present experiment. Another factor which was minimized was the extent to which a rational choice had to be made, given that a subject had already decided whether to work forward or backward on a particular occasion.

Berlyne (1965, p. 344) related working backward to Deutsch's (1960) theory of animal behaviour and to the theory of cue-producing responses derived from Hull (1935); but the results of animal experiments, whether positive or negative, seem to be consistent with the subject working, in thought, either forward or backward (cf. Anthony, 1964). Moreover, 'working backward' in the animal context cannot have the same reference to observation as it has in the human context, because the animal's overt behaviour can only be observed from the point where it is started in the problem situation. In the context of human problem-solving, however, 'working forward' and 'working backward' are descriptive terms, referring to a process which may be observed, or inferred directly from report.

This experiment was done at the University of Auckland, New Zealand, in 1964. A previous experiment, yielding broadly similar results, was done at the same university.

REFERENCES

ANTHONY, W. S. (1964). A Hullian reasoning situation. *Quart. J. exp. Psychol.* **16**, 378–82.
ARISTOTLE (1947). *The Ethics of Aristotle.* London: Dent.
BARTLETT, F. C. (1958). *Thinking.* New York: Basic Books.
BERLYNE, D. E. (1965). *Structure and Direction in Thinking.* New York: Wiley.
BRUNER, J. S., GOODNOW, J. J. & AUSTIN, G. A. (1958). *A Study of Thinking.* New York: Wiley.
DEUTSCH, J. A. (1960). *The Structural Basis of Behaviour.* London: Cambridge University Press.
DOLLARD, J. & MILLER, N. E. (1950). *Personality and Psychotherapy.* New York: McGraw-Hill.
DUNCKER, K. (1945). On problem-solving. *Psychol. Monogr.* 58, 270.
GEER, J. P. VAN DE (1957). *A Psychological Study of Problem Solving.* Haarlem: de Toorts.
HEATH, T. (1949). *Mathematics in Aristotle.* Oxford: Clarendon.
HEATH, T. L. (1908). *The Thirteen Books of Euclid's Elements*, vol. I. Cambridge University Press.

HULL, C. L. (1935). The mechanism of the assembly of behaviour segments in novel combinations suitable for problem solution. *Psychol. Rev.* **42**, 219–45.

MAIER, N. R. F. (1930). Reasoning in humans. I. On direction. *J. comp. Psychol.* **10**, 115–43.

MERRIFIELD, P. R., GUILFORD, J. P., CHRISTENSEN, P. R. & FRICK, J. W. (1960). A factor-analytic study of problem-solving abilities. *Rep. psychol. Lab.* no. 22. Los Angeles: University of Southern California.

MILLER, N. E. (1959). Liberalization of basic S–R concepts: extensions to conflict behaviour, motivation and social learning. In *Psychology: a Study of a Science*, vol 2. Ed. S. Koch. New York: McGraw-Hill.

NEWELL, A., SHAW, J. C. & SIMON, H. A. (1958). Elements of a theory of human problem-solving. *Psychol. Rev.* **65**, 151–66.

NEWELL, A., SHAW, J. C. & SIMON, H. A. (1962). The processes of creative thinking. In *Contemporary Approaches to Creative Thinking*, ed. H. E. Gruber, G. Terrell and M. Wertheimer. New York: Atherton.

POLYA, G. (1945). *How to Solve It.* Princeton University Press.

WOODWORTH, R. S. & SCHLOSBERG, H. (1955). *Experimental Psychology*. London: Methuen.

(*Manuscript received* 29 *October* 1964; *revised manuscript received* 25 *August* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 61–69
Printed in Great Britain

THE INFERENTIAL BASIS OF CONCEPT ATTAINMENT

By N. E. WETHERICK*

Medical Research Council Unit for Research on Occupational Aspects of Ageing, University of Liverpool

Three experiments are reported which investigated subjects' ability to make the kind of inference required in deriving a concept from positive or mixed positive and negative instances. Ability to make such inferences in a simple letter-group situation (and without the application of time pressure) was found to correlate highly with non-verbal intelligence. Three possible types of error are distinguished and it is shown that liability to one type increased with age even though the different age-groups were matched on non-verbal intelligence. An interpretation is suggested for this age difference.

The procedure followed in the experimental investigation of concept attainment usually involves presentation to the subject of a series of instances which vary along a few well-defined dimensions. The material used by Bruner is representative (Bruner, Goodnow & Austin, 1956); their stimulus cards varied in the kind of shapes on the cards, the number of shapes, the colour of the shapes and the number of borders. Each of these dimensions took one of three unequivocal values; for example, the colour was green, red or black. Some of the instances in the series exemplify the concept and some do not; a concept being defined as some particular value of one or more of the dimensions along which the instances may vary. For example, 'red', 'red circles', 'two red circles' could all have been concepts in Bruner's array. The series may be presented successively or simultaneously and the subject may be allowed to choose his own instances or be presented with instances chosen by the experimenter, but in any case his task is to extract from the instances the set of one or more attributes that characterizes positive instances (exemplifying the concept) but not negative instances. Where the dimensions of variation are unequivocal the result can be obtained by a process of elimination. Sets of one or more attributes that are present in both positive and negative instances or that are not present in every positive instance are irrelevant and may be eliminated. Because the number of possible sets of attributes is small the concept may be established by eliminating all but one of them. Bruner points out that the procedure he calls 'focusing' (in which dimensions are systematically tested for relevance) permits the concept to be established in as many trials as there are dimensions of variation but relatively few subjects are found to adopt this procedure.

The object of the present series of three experiments was to investigate the efficiency of the procedure by which sets of attributes are eliminated using positive or mixed positive and negative instances of different degrees of complexity. In the experiments visual material like Bruner's could have been used but for convenience of presentation letter-groups were adopted instead. In these letter-groups each position in the group constitutes a dimension of variation and the actual letters constitute values of attributes (A, B or C in the first position; D, E or F in the second and so on). The concept here is a letter or pair of letters that appears in all the positive instances (called 'Cat. I') but not in any negative instance ('Cat. II'). The headings 'Cat. I' and 'Cat. II' were used instead of 'Positive' and 'Negative' in an attempt to persuade subjects to pay more attention to 'negative' evidence than they would otherwise have been likely to do. The significance of membership of the two categories was defined operationally during the training procedures which preceded each experiment.

---

* Now Staff Tutor in Psychology, Department of Adult Education and Extra-Mural Studies, University of Liverpool.

METHODS

The tasks

In each of the three experiments the subject's task was to select a letter or a pair of letters that was in every positive instance presented but was not in any negative instance. The instances invariably eliminated all but one letter or pair of letters so that each problem had an unequivocal solution.

In Exp. I and Expt. II eight types of problem were presented. In the first four types, groups of three letters were used and the answer required was a single letter.

Type 1. Two positive letter-groups (CDH, ADI)—answer 'D'.

Type 2. Three positive letter-groups (BFH, BDH, AFH)—answer 'H'.

Type 3. One positive letter-group (ADI) and one negative (ADH)—answer 'I'.

Type 4. One positive letter-group (BFG) and two negative (AFI, BDI)—answer 'G'.

In the remaining types of problem, groups of four letters were used and the answer required was a pair of letters (any two of the four letters in the group constituting a pair):

Type 5. Two positive letter-groups (ADGL, BDHL)—Answer 'DL'.

Type 6. Three positive letter-groups (CEIK, CEIL, CEGK)—answer 'CE'.

Type 7. One positive letter-group (AFHK) and two negative (AEHK, AFHJ)—answer 'FK'.

Type 8. Two positive letter-groups (AEGL, CEGL) and two negative (BEIL, BFGL)—answer 'EG'.

In Expt. III twelve new types of problem were presented. All of them employed groups of four letters, but in six the answer required was a single letter and in the other six a pair of letters. The number of letter-groups involved in each, problem varied from two to five; the problems were otherwise similar to those already described.

Types of error

The structure of these problems makes it possible to distinguish three types of error. When a subject attempts to solve a concept attainment problem he has to consider a number of instances which may be either positive (exemplifying the concept) or mixed positive and negative. If the instances are all positive the appropriate operation is to eliminate those sets of attributes which are not present in every instance under consideration. If only one set remains the problem is solved, otherwise further instances are required. If the instances are mixed positive and negative the appropriate operation is to eliminate those sets of attributes which are present in both positive and negative instances, as well as those sets which are not present in every positive instance. If there remains only one set characterizing the positive instances the problem is solved, otherwise further instances are required. In performing these operations the subject may make three different types of error.

Type I. The subject may apply the wrong operation; that is, he may apply the operation appropriate to positive instances to mixed instances or vice versa. In the example given of problem type 7, FK must be the concept because it is the only pair of letters that is in AFHK but not in AEHK or AFHJ. A subject who applies the wrong operation will get the answer AH since this is the only pair of letters that is in all three. In the example given of problem type 6, CE must be the concept because it is the only pair of letters in CEIK and CEIL and CEGK. A subject applying the wrong operation might get the answer IK or IL or GK because each of these pairs of letters appears in one instance but not in the other two.

Type II. The subject may fail to take account of all the instances presented. In the example given of problem type 6 a subject making a type II error might get the answer CK or CI or EK since these pairs of letters are in two of the three instances. In the example given of problem type 7 the subject might get the answer AF or AK since these pairs of letters are in the positive instance and not in one of the two negative instances.

Type III. The subject may reverse the sign of his instances and choose a letter or pair of letters that is in one or more negative instances but not in any positive instance. In the example given of problem type 7 a subject making this type of error might choose EH or FJ.

General procedure

In all three experiments an elaborate training procedure was employed. The subject was taken step-by-step through worked examples which he kept by him throughout, together with a copy of the instructions. Questions could be asked at any time and were answered in full even if this meant a virtual repetition of the instructions. The instructions emphasized that there was no time limit and subjects appeared to have been convinced that this was the case. Though time scores might have been valuable none was taken because it is, in the writer's view, impossible to convince the ordinary adult subject that he is not required to hurry if he can see that he is being timed.

Subjects

Subjects were drawn from the Merseyside Volunteer Panel assembled by the M.R.C. Unit for Research on Occupational Aspects of Ageing. They were all male and between 20 and 75 years of age. Since a great deal of information was already available about the panel it was possible to select groups of subjects of different ages who were nevertheless matched for non-verbal intelligence. The test used was Raven's Progressive Matrices (1938).

EXPERIMENT I

Method

Procedure

This was a pilot experiment in which subjects were tested individually. The problems were presented one at a time on cards approximately 8 in. x 5 in. and the subject responded orally. Two series of problems were presented: series I consisted of twelve problems (three each of types 1–4 in random order); series II also consisted of twelve problems (three each of types 5–8 in random order).

Experimental plan

Three groups of ten subjects were employed; in the young group the subjects were between 21 and 30 years of age and had an average matrices score of 46·8 (range 38–53), in the middle-aged group they were between 42 and 49 years of age with an average matrices score of 46·5 (range 39–52) and in the old group they were between 60 and 70 years of age with an average matrices score of 46·5 (range 40–52).

Results

Table 1 shows the results; a preliminary report on these results was given by Wetherick (1962). The effect of matching different age-groups on non-verbal intelligence appears to have been to match them roughly on total number of correct

Table 1. *Expt. I: total scores on letter-group series I and II for three groups of subjects of increasing age (n = 10/group)*

Responses	Young group (20–30 yr)	Middle-aged group (40–50 yr)	Old group (60–70 yr)
Correct	149	145	139
Type I error	31	38	63
Type II error	37	32	20
Type III error	0	5	5
Unclassifiable error	0	3	3
'Insoluble'	23	17	10
Totals	240	240	240

responses to this test as well. This suggests the possibility of a relation between individual scores on this test and on the progressive matrices. A rank order correlation (Spearman's rho) was calculated but was not significant ($\rho = +0{\cdot}06$). From Table 1 it appears that the incidence of type I error arises with age and that the incidence of type II error falls, despite the fact that the age-groups were matched on non-verbal intelligence. Unfortunately Jonkheere's test for trend does not give a significant result in either case. Few unclassifiable errors occurred but subjects in this experiment were allowed to declare a problem 'insoluble' (though this was never the case). Old subjects appear to have been less willing to make use of this response than middle-aged or young.

EXPERIMENT II

Method

Procedure

In this experiment the subjects were tested in small groups. A proforma was used to present the problems in one of two conditions as follows:

	Cat. I	Cat. II	
Condition A	CDH, ADI	—	......
	BFG	AFI, BDI	......
Condition B	In CDH, ADI	—	......
	In BFG	Not in AFI, BDI	......

The subject recorded his response in the empty space on the right-hand side of the page. Four series of problems were presented: series I consisted of 24 problems (12 each of types 3 and 4 in random order), series II consisted of 24 problems (12 each of types 1 and 2 in random order), series III consisted of 48 problems (12 each of types 1, 2, 3, and 4 in random order), series IV consisted of 24 problems (6 each of types 5 to 8 in random order).

Experimental plan

Fifty subjects were used: two groups of five were drawn from each of the five decade age-groups 20–29, 30–39,..., 60–69. The ten groups so obtained were matched on non-verbal intelligence. No individual subject fell outside the range 39–49 and the group average varied only between 43 and 45 (no two groups from the same decade age-group varied by more than one point). One group from each decade age-group did the problems in condition A and the other in condition B. The two conditions differed only in the form of presentation adopted. The condition B group in the decade 20–29 had to be left one short as no suitable subject could be found to fill the gap. For this reason results are reported for forty-nine subjects only. None of them had taken part in Expt. I.

Results

Table 2 shows the results; a preliminary report on these results was given by Wetherick (1963).

The results of Expt. I had suggested that the number of correct responses might be related to non-verbal intelligence. For each condition in Expt. II a rank order correlation was again calculated between correct responses and scores on the Progressive Matrices. Both correlations were now significantly different from zero (condition A: $\rho = +0{\cdot}61$, $P < 0{\cdot}01$; condition B: $\rho = +0{\cdot}43$, $P < 0{\cdot}05$). The results of Expt. I also suggested that the incidence of type I error might increase with age even though subjects were matched for non-verbal intelligence. Expt. II was designed to investigate this possibility with still closer matching of subjects for age and intelligence, and to test a further hypothesis derived from the account of the genesis of type I errors already given. In a concept attainment problem the subject has to remember what the previous instances were and whether they were positive or negative. In Expt. II the instances were presented simultaneously (thus reducing the effort required to remember what the instances were), but in condition A the subject still had to remember which instances were positive and which were negative, the information being conveyed by their presence in either Cat. I or Cat. II. In condition B this effort was not required as the information was given directly for each problem 'In... Not in...'. Condition B should therefore reduce the incidence of type I error.

Table 2. *Expt. II: total scores on letter-group series I–IV for five groups of subjects of increasing age under each of two conditions, A and B (see text); n = 5/group, except for the* 20–29 *yr group of B for which n* = 4

	Condition A					Condition B					
Age-group (yr)...	20–29	30–39	40–49	50–59	60–69	20–29*		30–39	40–49	50–59	60–69
n...	5	5	5	5	5	4	(5)	5	5	5	5
Correct	527	551	487	517	482	465	(581·25)	584	565	559	585
Type I error	41	24	51	55	95	5	(6·25)	2	11	15	5
Type II error	28	14	18	20	12	8	(10·00)	7	7	8	8
Type III error	0	11	42	3	5	0	(0·00)	1	9	14	2
Unclassifiable error	4	0	2	5	6	2	(2·50)	6	8	4	0
Totals	600	600	600	600	600	480	(600·00)	600	600	600	600

* Estimated scores for *n* = 5 are given in brackets for comparison with the results for the other groups.

From Table 2 it appears that in condition A the incidence of type I error did in fact increase with age; Jonkheere's test for trend gives $Z = 2{\cdot}10$ ($P < 0{\cdot}025$) despite the fact that the highest individual type I error score was in the youngest age-group (one subject accounted for 28 of the 41 type I errors in this group and for 9 of the type II errors). In Table 2, condition B, the incidence of type I error is obviously very much lower than in condition A and it is no longer related to age.

Expt. II appears to have achieved its objectives. However, there was no relation in either condition between age and incidence of type II error.

EXPERIMENT III

Expt. III fell into two parts.

(1) The object in the first part was to test the generality of the results obtained in the previous experiments using different series of problems.

Procedure

Forty-two subjects were employed who did two new series of letter-group problems. Series I consisted of forty-eight problems presenting groups of four letters in which the answer was a single letter; series II consisted of thirty-six similar problems in which the answer was a pair of letters. The problems in both series were presented in Condition A.

Results

Table 3 shows the results obtained. The pattern of incidence of different types of error is similar to the one apparent in the previous experiments though in this case the age-groups were not matched on non-verbal intelligence. A rank-order correlation was calculated between scores on the series and on Raven's Progressive Matrices and was again significant ($\rho = 0{\cdot}66$, $P < 0{\cdot}001$).

(2) One of the principal advantages of the letter-group series is that equivalent series may be multiplied indefinitely by transforming the letter-groups in any regular way. The fact that the resulting series are equivalent is not at all apparent to the subject, and use was made of this in the second part of the experiment.

Table 3. *Expt. III: results on letter-group series I and II under condition A (see text) for three groups of subjects of increasing age*

	Young group (20–35 yr) ($n = 12$)		Middle-aged group (40–55 yr) ($n = 15$)		Old group (60–75 yr) ($n = 15$)		All ($n = 42$)	
Responses	Totals	Av.	Totals	Av.	Totals	Av.	Totals	Av.
Correct	937 (1171)*	78·1	1117	74·5	1086	72·4	3140	74·8
Type I error	50 (63)*	4·2	87	5·8	113	7·5	250	6·0
Type II error	21 (26)*	1·8	41	2·7	41	2·7	103	2·5
Type III error	0	0·0	15	1·0	20	1·3	35	0·8
Unclassified error	0	0·0	0	0·0	0	0·0	0	0·0

* The figures in brackets, given for comparison, are the actual scores of the young group increased by 25% to compensate for the smaller number of subjects in the group.

Procedure

Two sub-groups of fourteen subjects (matched for age and non-verbal intelligence) were selected from the original group of forty-two. A series of problems was constructed which was equivalent to series II by moving each letter one step through the alphabet (ABC... becoming BCD... and so on) and, after an interval of about twenty minutes (which included tea and a repeated digit span experiment which is not reported here), the two sub-groups did this equivalent series (series III)—one in Condition A and the other in Condition B.

Table 4. *Expt. III: comparative results of two sub-groups of fourteen subjects on letter-group series II (done by both groups in Condition A) and letter-group series III (done by the first group in condition A and by the second in condition B): series II and series III were equivalent forms*

	First sub-group		Second sub-group	
	Series II (Condition A)	Series III (Condition A)	Series II (Condition A)	Series III (Condition B)
Correct	473	485	370	470
Type I error	12	8	78	3
Type II error	19	10	43	31
Type III error	0	1	13	0
Unclassified error	0	0	0	0

Results

The comparative results obtained in series II (condition A) and series III (condition A or condition B) are shown in Table 4. The second sub-group turned out to be much more prone to error than the first but fortunately this does not effect the validity of the result. The first sub-group did what was in effect the same letter-group series twice within half-an-hour. As was to be expected the incidence of error decreased on the second trial but performance on the two series was significantly correlated ($\rho = 0{\cdot}84$, $P < 0{\cdot}001$). The two equivalent forms ranked subjects in virtually the same order of proficiency. The second sub-group did the series first in condition A and then in condition B. The incidence of Type I error decreased from 78 in condition A to 3 in condition B. This supports the finding obtained in Experiment II and in a more powerful case since here the same subjects were employed in both conditions. Even though condition B virtually eliminates Type I error, performance on the two series was again significantly correlated ($\rho = 0{\cdot}50$, $P < 0{\cdot}05$).

DISCUSSION

It is very often assumed that in his work on concept attainment Bruner was concerned with a class of 'categorization' problems of only limited interest for the study of thinking generally. It is not difficult to show that Bruner himself does not take this view, he writes: 'It has been our intention to investigate paradigms of behaviour that represent in stripped-down form the kind of behaviour called problem-solving or thinking in everyday life', nor is it difficult to show that a great deal of normal thinking (in Bruner's view 'virtually all') is 'dependent on the process of categorizing.' It follows that in real life we must continually be called upon to make inferences like the ones required in the three experiments reported here, though the inferences will not normally be so simple and unambiguous. The importance in real life of the ability to make such inferences may account for the high correlations obtained in Expt. II and Expt. III between scores on the letter-group problem series and non-verbal intelligence. (The failure to obtain a correlation in Expt. I may have been due to the small number of problems that was used: twenty-nine of the thirty subjects in the experiment scored between 11 and 19 out of 24, and the test may have been too short to be sensitive to small differences in ability.) The problem series test the ability to perform a task of very great apparent simplicity. All that the subject has to do is to consider three or four groups of letters and to find a letter or a pair of letters that is in all of them, or that is in one or two of them, but that is not in the rest. This task is performed in the absence of time or any other kind of pressure, with instructions and worked examples available for reference. By comparison some of the items in Raven's Matrices are difficult enough in themselves to extend subjects of a high level of mental ability and the instructions are complex and given only once.

The structure of the problems employed in these experiments makes it possible to distinguish different types of error. The incidence of type I error has been shown to increase significantly with age even though the age-groups are matched on non-verbal intelligence. The increased incidence of this type of error is however compensated in part by a decreased incidence of other types of error to the extent that no tendency was apparent for total error to increase with age. It appears that there may be a difference in kind between the measured non-verbal intelligence of old and young persons as well as the commonly observed difference in degree. Welford (1958) has considered whether the apparent 'rigidity' of the older person arises from an increasing unwillingness to modify established opinions or from an increasing inability to do so and concludes in favour of the latter hypothesis. Increasing liability to type I error may be one of the factors that contributes to this increasing inability. Consider the case of a person who expresses the prejudiced opinion that 'All Negroes are dirty'. This implies a belief that all instances having the attribute 'being a Negro' are positive exemplars of the concept 'dirty'; that is, that the observation of an instance exemplifying the attribute 'being a Negro' enables the observer to predict that the instance will also exemplify the attribute 'dirty'. To confute such an opinion it is necessary to produce negative instances of the concept 'dirty' which have the attribute 'being a Negro'. 'Being a Negro' ought then to be eliminated as an attribute predicting 'dirty' on the ground that it characterizes both positive and negative instances of the concept. If however the person expressing the opinion is liable to type I error the new evidence may have exactly the same effect on him as if the instances had been positive instead of negative. The difficulty that an old person may encounter in trying to deal appropriately with positive and negative instances of a concept is emphasized by the finding in Expt. II that presentation in condition B not only reduced the overall incidence of type I error, but also abolished its relation with age. The benefit received by the old person is proportionately much greater.

Liability to type II error may be a consequence of limited capacity of the short-term memory store. Welford has suggested that limited short-term memory capacity may prevent a subject solving a given problem because finding the solution demands the more or less simultaneous consideration of a number of different pieces of information and the subject's capacity is simply not large enough to accommodate them all. The best solution that such a subject can hope to achieve may be one based on the largest number of pieces of information he is capable of considering simultaneously. Letter-groups may be regarded as 'pieces of information' in this sense and the subject who chooses a letter that is in two out of three letter-groups when he is looking for a letter that is in all three seems to be making the kind of error Welford has in mind. An age-related decline in the capacity of the short-term store has been observed in a variety of experimental stituations which might lead us to expect an age-related increase in the incidence of type II error. No such relation was established in these experiments: in Expt. I there was what appeared to be an age-related decrease, but this result was not verified in the later experiments. It may be, however, that the matching procedure that was adopted masked an increased liability to type II error which would otherwise have been apparent in the older subjects. Not many previous investigators have been in a position to match their age-groups on non-verbal intelligence.

Little has been said about type III errors. The incidence of errors of this type was low. They appear to occur mainly in the middle-aged, but the relation cannot be demonstrated statistically. What is perhaps most notable about them is that they do sometimes occur. The possibility of their occurrence follows from the theoretical analysis proposed in this paper and would not otherwise have been likely to be predicted.

In this paper some of the findings from three experiments have been discussed. More detail together with some consideration of the theoretical issues involved is to be found in Wetherick (1965).

I am indebted to Dr Alastair Heron, Director of the M.R.C. Unit for Research on Occupational Aspects of Ageing at the time this work was done (now Director of the Institute for Social Research, Lusaka, Zambia), to my colleagues at the Unit and to Dr D. B. Bromley of the Department of Psychology, Liverpool University, for advice and assistance throughout.

REFERENCES

BRUNER, J. S., GOODNOW, J. J. & AUSTIN, G. A. (1956). *A Study of Thinking*. New York: Wiley.

RAVEN, J. C. (1938). *Progressive Matrices, Sets A, B, C, D and E*. London: H. K. Lewis.

WELFORD, A. T. (1958). *Ageing and Human Skill*. London: Oxford University Press.

WETHERICK, N. E. (1962). Age related differences in patterns of inference in a non-specific task. *Bull. Brit. psychol. Soc.* no., 47, 34A.

WETHERICK, N. E. (1963). Patterns of error in inductive reasoning. *Bull. Brit. psychol. Soc.* **16**, no. 51, 31A.

WETHERICK, N. E. (1965). Problem-Solving Ability and Age in the Normal Human Adult. Unpublished Ph.D. Thesis, University of Liverpool.

(*Manuscript received* 2 *September* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 71–75
Printed in Great Britain

THE RELATION BETWEEN BREADTH OF CATEGORY AND DECISION TIME

BY HENRI TAJFEL
University of Oxford

AND JEROME S. BRUNER
Harvard University

Twenty-four subjects were presented with a series of stimuli, all of equal length (5 in.), and asked to classify each as being or not being 5 in. long. The subjects were then divided into three groups (broad, middle and narrow categorizers) according to the number of including and excluding responses that they gave.

It was found that: (*a*) the decision times of both extreme groups were considerably shorter than those of the middle group; (*b*) there was a tendency for the extreme groups to have shorter decision times for their preferred type of response; but this tendency did not reach statistical significance.

Bruner & Tajfel (1961) reported a series of experiments concerned with cognitive behaviour related to breadth of category. They found some clearly consistent differences in the type of 'cognitive risk' taken by two extreme groups of subjects: the 'broad' and the 'narrow' categorizers. The broad categorizers, when faced in an ambiguous judgement situation with the task of deciding whether an array of stimuli belongs or does not belong to a previously defined inclusion class, assign to that class a large proportion of the stimuli. The narrow categorizers exclude a large proportion of stimuli from the target class. The tendency to use broad or narrow inclusion classes was found to be consistent for an individual, independently of the nature and the distribution of the stimuli presented to him.

In these experiments children aged about eleven were used as subjects. The results were subsequently confirmed with groups of adult subjects, though individual consistencies of judgement were not as marked for adults as they were for children (Tajfel, Richardson & Everstine, 1964*a*). In a further study (Tajfel, Richardson & Everstine, 1964*b*) it was found that these more moderate consistencies were maintained by adult subjects even when a system of rewards and penalties for various types of incorrect assignments was introduced.

The present experiment was conducted in order to explore a collateral aspect of this phenomenon. If the subjects who find themselves consistently at one or the other extreme of the distribution of responses do respond according to their consistent mode of resolving doubt rather than in terms of requirements imposed by the nature of the stimuli, then it may be expected that they reach their decisions without engaging in a close inspection of the stimuli. This should be reflected in over-all shorter decision times for those subjects than for the 'middle' group, which is not characterized by the predominant use of one type of response.

A second prediction follows from the findings reported above. Bruner & Tajfel (1961) found that there were no differences between the two extreme groups in their capacity to discriminate between the stimuli; their performance was determined to a large extent by internal requirements rather than by the frequency and distribution of the stimuli presented to them. Therefore, there should be a difference between them in the relative ease with which they make an 'including' or an 'excluding' decision. From this follows the expectation that consistent differences should be found in the amount of time that subjects in these groups need in order to make these two kinds of decisions. Thus, in the group of broad categorizers decisions to include should be made in shorter time than decision to exclude; and the converse for the narrow categorizers.

METHOD

Procedure

The first requirement for testing these predictions was to isolate the decision time variable in a 'pure' form by making it independent of the differences between the stimuli presented to the subjects. This was done by using a series of stimuli which did not differ from one another in the dimension along which judgements were made. As in Expt. III of Bruner & Tajfel (1961), 'stimuli consisted of straight lines drawn on rectangles of white cardboard, 11 in. by 14 in. in size. The inclinations and positions of lines on the rectangles varied randomly' (p. 239). There was one line to each rectangle, forty-eight in all. In contrast to the previously reported experiments, all the lines were of the same length, 5 in. 'The subjects, tested individually, were seated at one end of a table, the experimenter at the other, separated by a distance of about 3 ft.' (Bruner & Tajfel, 1961, p. 239). The lines were shown one by one, the order of presentation being the same for all the subjects. They were told to decide whether each line was 5 in. long or not. If they thought it was, they were to say 'yes'; if they thought not, 'no'. Two lines were shown first as specimens of 5 in. length. The presentation of the experimental series followed immediately. Time exposure was not limited, the cardboard rectangle being held up until the subject made his decision.

Response measures

Two scores were taken for each subject: the frequency of his 'yes' and 'no' responses, and for each response the decision time from the presentation of the stimulus until the subject made his decision.

Subjects

Twenty-four children aged between eleven and twelve who had not taken part in previous experiments on breadth of category served as subjects.

RESULTS

In the scoring of frequency and decision times, the first four stimuli were not taken into account, as some preparation was needed for the children to settle down and familiarize themselves with the task. For the remaining 44 stimuli, the range of 'yes' and 'no' responses was from 39 to 5 for the 'broadest' subject to 14 to 30 for the 'narrowest'.

The division of the twenty-four subjects into three groups of broad, middle and narrow categorizers was made at cut-off points nearest to 27 % of the distribution, as in the previous experiments. This resulted in five subjects with the highest proportion of 'yes' responses being assigned to the broad group (because of a run of ties after the fifth subject), and six subjects with the lowest proportion of 'yes' responses to the narrow group. The middle group consisted of the remaining thirteen subjects. The range for the broad group was 39–34 'yes' responses; for the middle group 30–22 'yes' responses; for the narrow group 21–14 'yes' responses. The mean number of 'yes' responses was 36 for the broad group; 26 for the middle group; and 18 for the narrow group.

(*a*) *Over-all decision times*

The first prediction was clearly confirmed. Table 1 shows that both extreme groups tend to have considerably shorter decision times than the middle group. This applies both to 'yes' and to 'no' responses. There is practically no difference in the over-all decision times of both extreme groups. The performance of each of them taken separately is significantly different from the performance of the middle group

Table 1. *Mean decision times* (*in sec*) *for a group of children* (11–12 *yr*) *divided into broad, middle and narrow categorizers*

Group	n	'Yes' responses	'No' responses	All responses
Broad	5	2·00	2·32	2·08
Middle	13	2·94	2·96	2·95
Narrow	6	2·22	2·08	2·13
All groups	24	2·56	2·61	2·58

($P = 0{\cdot}05$ for the difference between the broad and the middle groups; $P < 0{\cdot}05$ for the difference between the narrow and the middle groups, as assessed by Mann–Whitney U tests). The difference between both extreme groups combined and the middle group is significant ($0{\cdot}05 > P > 0{\cdot}01$).

In previous experiments the distribution of inclusion and exclusion decisions remained relatively constant for 'extreme' subjects in situations involving widely different distributions of stimuli. It was as if they had made up their minds in advance and then took their decision on the basis of a negligible cue. This reflected itself in relatively short decision times for these subjects in the case of *both* kinds of response. It is interesting that this result was obtained with the distribution of stimuli used in the present experiment. Objectively, *all* the stimuli fell into the inclusion class; therefore it could be expected that, if the subjects from all three groups did not differ in their manner of inspecting the stimuli, the group af narrow categorizers would show the greatest hesitation in adopting their preferred mode of response, the broad categorizers would find the task easy and congenial, and the middle group would lie between these extremes. And yet, the pattern of decision times shown by the three groups is not related to their respective frequencies of including responses. The similarity of the two extreme groups in their method of approaching the task leads to a close resemblance between them in their decision times despite the drastic reversal of the relation between the objective requirements of the situation and the distribution of responses that they respectively exhibit.

(*b*) *Decision times for 'yes' and 'no' responses*

The second prediction was that broad categorizers would tend to have shorter decision times for their 'yes' than for their 'no' responses, and conversely for the narrow categorizers. Table 1 shows that there is a tendency in the expected direction: (*a*) there is no difference between the mean decision times for 'yes' and 'no' responses in the middle group; (*b*) in the broad group, the decision times for 'yes' responses tend to be shorter than for 'no' responses; (*c*) in the narrow group, the converse is true; (*d*) the mean decision times of 'preferred' responses for both groups together (i.e. 'yes' for the broad group and 'no' for the narrow group) is 2·04 sec as compared with the mean decision times of 2·26 sec for 'non-preferred' responses (i.e. 'no' for the broad group and 'yes' for the narrow group).

The prediction, however, does not find reliable support in the data. The statistical significance of the differences in the subjects' scores between decision times for their 'yes' and 'no' responses was assessed using the Wilcoxon matched-pairs signed-ranks test. The hypothesis was tested for the groups of broad and narrow categorizers, and also by combining the two groups with regard to the decision times for their 'preferred' responses. The differences between the subjects' scores for the two types of responses did not reach statistical significance either for the separate groups or for the two groups combined. It was also found that the respective decision times for 'yes' and 'no' responses did not differ significantly between the two extreme groups of subjects.

DISCUSSION

The general implications of the present study cannot be considered without reference to the previous work briefly summarized in the introduction. The present data do not, by themselves, provide evidence that our groups of 'extreme' subjects were genuinely broad and narrow categorizers, nor do they allow for inferences concerning possible personality differences between these two groups.

The problem of the relation of personality differences to individual modes of categorizing is a complex one, and the evidence contains many inconsistencies. These inconsistencies (cf. Tajfel *et al.* 1964*a*, for a review) can sometimes be attributed to differences in the nature of the tasks employed, sometimes to other situational variables, and sometimes to difficulties in identifying those personality variables which could clearly and explicitly be related to the type of responses requested from the subjects. The aim of the present study was not to attempt a further elucidation of these problems but rather to provide supplementary and independent evidence that the individual differences in categorizing discovered in previous studies are based on stable tendencies displayed by individuals to use one or another mode of response in their handling of ambiguous judgement decisions of a certain kind. At the same time, the differences in decision times reported here reveal something of the manner in which these decisions are reached.

The results of the experiment can be considered to provide such supplementary information only if there are reasons to believe that the individual differences in the frequency of including and excluding responses exhibited by the subjects do reflect more general differences in breadth of categorizing. The assumption that this was the case is based on results from previous work (Bruner & Tajfel, 1961; Tajfel *et al.* 1964*a*, *b*) in which the subjects' performance in the task used in the present study was found to correlate with performance on other judgement tasks. An alternative explanation of individual differences shown by the subjects in the present study could be formulated in terms of differences in constant error in the judgements of the length of the lines. But such an explanation would not account for the differences in the decision times between the various groups of subjects, while the hypothesis relating these differences to preferred modes of categorizing finds a clear confirmation in the data.

This research was done as part of the Cognition Project at Harvard University and was supported in part by USPHS grant M-1324. Our thanks are due to Mr William Rohwer (at Harvard College at the time when the present study was conducted) for his help in running the experiment.

REFERENCES

BRUNER, J. S. & TAJFEL, H. (1961). Cognitive risk and environmental change. *J. abnorm. soc. Psychol.* **62**, 231–41.

TAJFEL, H., RICHARDSON, A. & EVERSTINE, L. (1964*a*). Individual consistencies in categorizing: a study of judgmental behavior. *J. Pers.* **32**, 90–108.

TAJFEL, H., RICHARDSON, A. & EVERSTINE, L. (1964*b*). Individual judgment consistencies in conditions of risk taking. *J. Pers.* **32**, 550–65.

(*Manuscript received* 18 *March* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 77–86
Printed in Great Britain

MEMORY AND DISTORTION OF MEANINGFUL WRITTEN MATERIAL*

BY ROBYN MASON DAWES
Ann Arbor VA Hospital, Michigan, U.S.A.

A method is presented for measuring memory and distortion of meaningful written material. The method is based on the fact that since meaningful material asserts set relations, a subject's memory and distortion of such material may be measured by asking him to recognize or recall set relations, rather than specific verbal units. In addition, a measure of 'simplification', in terms of distorted set relations, is proposed. A series of experiments concerned with recognition and recall of set relations reveals that simplification, as defined, does occur. But it does not increase over time.

In our literate society, a primary means of acquiring knowledge is reading. The memory of what we have read, and our distortion of it, is therefore a major determinant of what we know, think we know, and don't know. Systematic studies of memory of meaningful written passages have, however, all but disappeared from the psychological literature of the past three decades. Despite vigorous criticisms by Bartlett (1932, pp. 3–6), psychologists investigating memory of written materials have dealt almost exclusively with discrete linguistic units, usually single words or nonsense syllables. Moreover, those early authors who did investigate memory of meaningful written passages (e.g. Ballard, 1913; Gordon, 1925; Henderson, 1903; Williams, 1926) *measured* memory in terms of discrete linguistic units, i.e. in terms of their subjects' ability to recall single words or phrases. Bartlett is an exception. But while Bartlett studied more global aspects of his subjects' recall (such as its conceptual complexity), he had no way of measuring these aspects; he was forced to rely on qualitative judgements about his subjects' reproductions.

The aim of the present paper is to present a method for measuring memory of meaningful material and to review a series of studies making use of this method. This method is meant to enable its user to study broad aspects of a subject's memory, rather than his ability to remember discrete units; in particular, this method hopefully provides a measure of one of the aspects of memory most discussed by Bartlett—simplification.

The measurement method is based on the fact that *meaningful declarative statements assert set relations*; in particular, such statements assert a set relation between their subject class and their predicate class. (For example, the statement 'the moon is blue' asserts that the set consisting of the single object 'moon' is wholly contained in the set of 'blue things'. The central thesis of the present paper is that memory and distortion of meaningful material may be measured by the memory and distortion of the set relations asserted in that material. Set relations embody 'what is said' in a passage. At the same time, a given set relation may be expressed in a number of different ways, so that memory for set relations does not reduce to memory for individual verbal units—nor to memory for grammatical structure.

---

* All statements are those of the author and do not necessarily represent the opinions or policy of the Veterans Administration.

First, it is necessary to explicate a system for classifying set relations. The system presented here is that developed by the French mathematician J. D. Gergonne (1817). It distinguishes four types of set relations. (Gergonne actually enumerated five relations, not four, because he distinguished between the relation in which a set X is included in a set Y and the relation in which Y is included in X. This distinction is unnecessary for present purposes.) (*a*) The first relation two sets X and Y may have is that of *identity*: the relation in which all elements of X are elements of Y and vice versa. (*b*) The second relation two sets X and Y may have is that of *exclusion*: the relation in which X and Y have no elements in common. (*c*) The third relation two sets X and Y may have is that of *inclusion*: the relation in which all elements of one set (say X) are elements of the other but not all elements of the other (Y) are elements of the first (X). (*d*) The fourth relation two sets X and Y may have is that of *disjunction*: the relation in which X and Y have some elements in common but neither set is included in the other.

As Gergonne pointed out, *any two sets have one of these relations and only one*. Thus, an unambiguous declarative statement or group of statements specifies exactly one of these relations as existing between the sets with which the sentence or sentence group is concerned. It follows that memory of meaningful declarative material may be measured by asking the subject exposed to the material to specify which of the Gergonne relations holds between the various sets. If the subject specifies the correct relation, he has remembered (or forgotten and guessed correctly); if the subject cannot specify the relation, he has forgotten, and if he specifies the wrong relation he has distorted (or forgotten and guessed incorrectly).

Unfortunately, the English language is not as unambiguous as the Gergonne relations. For example, the statement with the structure 'some X are Y', e.g. 'some Greeks are mortal', is true when X and Y have an identity relation, when X is included in Y, when Y is included in X, or when X and Y have a disjunctive relation; thus, the statement does not assert one of the Gergonne relations. Such ambiguous statements do, however, eliminate certain of the Gergonne relations—otherwise they would have no meaning at all; for example, the 'some X are Y' statement eliminates the exclusion relation. In addition, ambiguous statements are often rendered unambiguous through convention; for example, we rarely (as above) state 'some X are Y' when it is true that 'all X are Y', and we rarely state 'all X are Y' when it is also true that X and Y are identical.

Thus, the modifier 'some' usually implies disjunction, not inclusion; the modifier 'all' usually implies inclusion, not identity; and so on. While these conventions are accepted in most writing, however, there are exceptions. In the material used in the series of studies to be reported here, ambiguity and consequent reliance on conventions are carefully avoided.

The relations labelled 'identity', 'exclusion', and 'inclusion' all have one property in common. One set is entirely *nested* in the other or in the complement of the other. When, for example, two sets have an inclusive relation, one is nested in the other; when two sets have an exclusive relation, each is nested in the complement of the other; finally, when two sets have an identity relation, they are mutually nested. If two sets have a disjunctive relation, on the other hand, neither is nested in the other or in the complement of the other.

This distinction between nested and disjunctive relations forms the basis for a measure of 'simplification'. It is proposed that prose material has been simplified whenever a disjunctive relation is remembered as nested, and it has been made more complex whenever a nested relation is remembered as disjunctive. The former type of distortion, that of remembering a disjunctive relation as nested, has been termed *overgeneralization*, while the latter type has been termed *pseudo-discrimination* (see Dawes, 1964: 'overgeneralization' is used here in this sense and none other; it needs to be distinguished from 'over-inclusion', which refers to a tendency to subsume too many objects in a given set; the latter term deals with a distorted relation between a set and set elements, while overgeneralization has to do with a distorted relation between two sets). It is proposed that 'simplification' may be measured by noting the relative amount of overgeneralization as opposed to pseudo-discrimination in memory of written material.

Another undefined but often used notion is that of 'structure'. A rigorous definition of 'structure' among sets is proposed: two sets have a 'structured' relation if and only if knowledge that an object is a member of one of these sets *may* yield knowledge of whether or not it is a member of the other.

It immediately follows that nested sets have a structured relation and disjunctive sets have an unstructured relation. If, for example, the set X is included in the set Y, knowledge that an object is a member of X yields knowledge that it is also a member of Y; moreover, knowledge that an object does not belong to Y yields knowledge that it does not belong to X. (Knowledge that an object belongs to Y yields no knowledge of whether it belongs to X; hence the term 'may' in the definition of 'structure'.) The reader may verify that the other nested relations are structured and that the disjunctive relation is not. It follows that *simplification of written material increases its structure.* Bartlett also maintained that simplification resulted in greater structure; in fact, he postulated that simplification was even *due to* a quest for structure—which he termed an 'effort after meaning' (1932, p. 227).

EXPERIMENT I

The purpose of this experiment was to test subjects' memory for the set relations asserted in two stories immediately after reading them and three days later. Two stories written by the author were used.

The stories

Both stories dealt with social conflicts between various political and economic groups, and the Gergonne relation existing between each pair of groups was made explicit in the story.

Story 1 was a fantasy about an island on which there are two main occupations, farming and ranching; no one on the island is both a farmer and a rancher. The ranchers are much more prosperous than the farmers, and all the island administrators are ranchers. The farmers propose the construction of an irrigation canal to allow the island's interior to be cultivated. Although this plan is opposed by most of the ranchers, the farmers persuade a few administrators to back it, and these administrators together with some farmers form a pro-canal association. The island is run democratically and this association forces the issue to a vote. All the islanders vote, with all the members of the pro-canal association and all the farmers voting for construction and everyone else voting against it. The majority vote in favour of construction, but the island's administrators decide it would be dangerous to have a canal more than two inches wide and three inches deep.

Hence, five groups of people are discussed in the story: farmers, ranchers, administrators, members of the pro-canal association, and people who voted in favour of the canal. The reader may verify for himself that half of the (ten) relations between these groups are nested and half disjunctive.

Story 2 was a similar story. It concerned a hypothetical situation in Northern Italy, in which there was a conflict over whether to subsidize farmers. The important groups in this story were cliques in the governing assembly: rural Social Democrats, rural Communists, urban Social Democrats, urban Communists, and supporters of a subsidy plan.

These stories were similar to those presented in Dawes (1964), though not identical with them. Copies of the stories discussed here may be obtained from the author on request.

Procedure

The stories were typed, in the order given above, in a booklet entitled 'Political Decision Survey (form A): National Opinion Survey Corporation'. The avowed purpose of this survey was explained in the instructions on the following page. These instructions read: 'The purpose of the survey is to study the political and economic decisions people reach when these decisions are based on such evidence as newspaper or magazine stories. In this booklet, you will find newspaper-like stories concerning hypothetical issues. *Your job is to read each story as you would a newspaper article in which you were interested* in order to reach a decision about how you would vote on the issue. Do not study or reread the stories, for we are interested in decisions based on information you obtain through your usual reading of a newspaper. If you wish to think over the issues, do so at the time you are asked to reach your decision about how you would vote.' Thus, the subjects were *not* instructed to memorize, but rather to comprehend, the material presented to them.

After reading both stories and reaching his decision about how he would have voted, each subject was asked questions of the following form.

'Circle the correct alternative: (*a*) No rancher voted for construction of the canal. (*b*) Not all but some ranchers voted for construction of the canal.' That is, each question consisted of one assertion of a disjunctive relation between two sets and one assertion of a nested relation, and the subject was to indicate which assertion was correct.

There were twelve questions in all, eight about story 1 and four about story 2. The disjunctive assertion was always worded 'not all but some . . . are . . .'. This wording was used to emphasize that as well as having some elements in common, the two sets had elements which were not in common. The nested assertion was worded either 'all . . . are . . .' or 'no . . . are . . .'; seven of the nested assertions were of the former type, five of the latter. As there are always two possible nested relations that sets may have, inclusion and exclusion, the nested assertion that made most sense in the context of the story was presented; there was never any difficulty in deciding which to present.

The statement used to assert a disjunctive alternative, 'not all but some X are Y', is also true of the nested relation in which Y is included in X. If such inclusion were possible in the context of the story, the disjunctive assertion was not worded 'not all but some X are Y', but 'not all but some Y are X'.

Since all sets have either a disjunctive relation or a nested one, each question had one correct and one incorrect assertion. If the correct assertion was the disjunctive one, the subject makes an overgeneralization if he circles the nested assertion; if the correct assertion is the nested one, the subject makes a pseudo-discrimination if he circles the disjunctive assertion. The correct assertion in six of the twelve questions was the nested one and in the other six the disjunctive one. Thus, subjects had an equal opportunity to make either overgeneralizations or pseudo-discriminations.

In addition to answering the questions immediately after reading the stories, subjects returned 3 days later and again answered these questions. This experiment was embedded in a number of other experiments, so that the subjects did not know they would be asked these questions a second time.

Subjects

Subjects were fifty-four male and female undergraduate students at the summer session of the University of Michigan. Previous work (Dawes, 1964, p. 454) has indicated that males and females do not differ in their memory for this type of material. Twelve other subjects were discarded: eight because English was not their first language, three because they failed to attend the second experimental session, and one because she omitted one question.

Results

The following notation was adopted in order to facilitate the analysis of results: N—number of correct nested assertions circled; N'—number of incorrect nested assertions circled (i.e. overgeneralizations); D—number of correct disjunctive assertions circled; D'—number of incorrect disjunctive assertions circled (i.e. pseudo-discriminations).

The first statistic of interest is the number of correctly circled assertions ($N + D$). This statistic is termed a c-score (or correct score). The mean c-score for the first testing is 8·09 (S.D. = 1·78). The number of correct answers expected by chance alone would be six: the mean c-score is reliably above this chance level ($t = 8{\cdot}36$; D.F. 53; $P < 0{\cdot}001$). The mean c-score on the testing 3 days later was 8·11 (S.D. = 1·87). In accord with the closeness of the two mean c-scores, the subjects' difference scores (between first testing c-score and later testing c-score) do not differ significantly from zero ($t = 0{\cdot}11$; D.F. 53; N.S.).

A second question of interest is the amount of overgeneralization relative to the amount of pseudo-discrimination, since this is our proposed measure of simplification. The relevant statistic is ($N' - D'$), termed an s-score (or simplification score). The mean s-score on the first testing is 1·65 (S.D. = 1·78). An s-score of zero would be expected by chance alone, since there are an equal number of questions in which the nested assertion is incorrect and in which the disjunctive assertion is incorrect. A t-test indicates that the mean s-score is reliably greater than zero ($t = 6{\cdot}72$; D.F. 53; $P < 0{\cdot}001$).

The mean s-score on the testing 3 days later was 1·04 (S.D. = 2·10). This s-score is also reliably greater than zero ($t = 3{\cdot}61$; D.F. 53; $P < 0{\cdot}001$). In addition, however, a t-test performed on the subjects' difference scores (between first testing s-score and second testing s-score) reveals that these differences are significantly *less* than zero ($t = 3{\cdot}50$; D.F. 53; $P < 0{\cdot}001$). The s-scores after the 3-day interval are smaller.

Discussion

The first finding is that recognition of set relations asserted in previously read material is more accurate than would be expected by chance alone, but, surprisingly, the accuracy of recognition 3 days later is just as great. Secondly, simplification, as measured by errors in the recognition of set relations, is greater than would be expected by chance alone, but, again surprisingly, simplification *decreases* over time, rather than increases.

Bartlett previously found that simplification increases over time; this finding is not only supported by his data but is also supported by everyday observation. The present results are in direct conflict with his. One possible explanation for this conflict is that the subject is less sure of his memory after the 3-day delay, and he has a response bias toward choosing the disjunctive assertion when he is uncertain; that is, the subject tends to choose 'some' rather than 'all' or 'none' when he is unsure. There is evidence that such a response bias may exist (Dawes, 1964), but the evidence is far from conclusive.

This possibility illustrates a problem in the current methodology; when the subject is asked to recognize set relations, forgetting and distortion are confounded. That is, the subject may circle the incorrect assertion either because he has distorted the material and now believes the incorrect assertion is correct, or because he has forgotten the material and happens to guess wrong. In addition, selective memory and distortion are confounded. This confounding is illustrated by the fact that, in order to measure the tendency to recognize nested sets, undoubtedly we should have considered the statistic $(N - D)$, the number of correctly recognized nested relations minus the number of correctly recognized disjunctive relations. But this statistic is the same s-score we have already considered, because $(N - D) = (6 - D') - (6 - N') = N' - D'$.

Thus, it would be desirable to run an experiment in which the subject was asked to recall, rather than recognize, set relations. It would also be desirable to find out how groups performed on the delayed testing without taking the first testing. Experiment II was designed to fill these needs.

EXPERIMENT II

Procedure

The same stories were used. There were, however, two different sets of questions about the stories. One, the recognition set, was the same as that employed in Expt. I. The other, the recall set, began with a list of the groups referred to in the recognition set; the subject was then asked to recall the set relations of these groups. The instructions read: 'First, list all the group relationships you remember that have the form "all . . . were . . ." or " no . . . were . . .", where the blanks refer to the groups. For example, in the first story: "All members of the procanal association were people who voted in favour of the canal."' After a blank sheet, a second question read: 'Now list all the group relationships you remember that have the form "not all but some . . . were . . .", where the blanks again refer to the groups. For example, in the first story: "Not all but some senators were people who voted in favour of the canal."' The order of the two questions was balanced throughout the experimental conditions described below.

Subjects

Sixty-four male and female undergraduate students attending the University of Michigan's spring semester read the two stories with the same instructions as those used in Expt. I. They met in three groups of ten or more. Since seven subjects failed to reappear for the second testing, a number of subjects had to be run individually in addition or in small groups in order that there be sixteen subjects in each experimental group.

Plan

The subjects were randomly divided into four equal groups, who received slightly different booklets. The groups were formed from two dichotomies; the first was between recognition and recall testing; the second was between subjects whose memory was tested both immediately after reading and also 2 days later and subjects who were only tested two days later. This design is shown in Table 1.

Table 1. *Experimental design for Expt. II*

	Immediate and delayed testing	Delayed testing only
Recognition testing	Group A ($n = 16$)	Group B ($n = 16$)
Recall testing	Group C ($n = 16$)	Group D ($n = 16$)

As before, the experiment was embedded in a number of others, so that the subjects did not know that they would be required to answer questions about the stories during the second experimental session.

Results

Table 2*a* gives mean *c*-scores $(N+D)$ for the two recognition groups. There is no significant difference between the immediate and delayed conditions for group A ($t = 0{\cdot}20$; D.F. 15; N.S.). The mean *c*-score for group B in the delayed condition is somewhat less than that for group A in the immediate condition ($t = 1{\cdot}96$; D.F. 30; $P < 0{\cdot}10$), and also than that for group A in the delayed condition ($t = 1{\cdot}97$; D.F. 30; $P < 0{\cdot}10$). The third of these t values is not of course independent of the other two. As in Expt. I therefore, the accuracy of the group tested both immediately and after a delay does not decrease; there is however a tendency for accuracy to be impaired in the group tested only after a delay.

Table 2. *Mean scores and standard deviation for recognition and recall in Expt. II*

	Immediate		Delayed	
Group	Mean	S.D.	Mean	S.D.
		(*a*) *c*-scores and S.D.		
A	8·38	1·42	8·44	1·64
B	—	—	6·94	2·43
		(*b*) *s*-scores and S.D.		
A	1·13	1·80	0·81	2·01
B	—	—	1·94	2·07
		(*c*) *c*-scores and S.D.		
C	4·25	2·25	4·50	2·50
D	—	—	4·13	3·21
		(*d*) *s*-scores and S.D.		
C	2·19	1·91	2·50	1·58
D	—	—	2·31	1·53
		(*e*) *a*-scores and S.D.		
C	0·06	3·60	1·13	3·44
D	—	—	−0·93	2·64
		(*f*) *s*′-scores and S.D.		
Successive	1·38	1·57	0·81	1·43
Single	—	—	−0·50	3·06

Table 2*b* gives mean *s*-scores $(N' - D')$ for the two groups. The mean *s*-score for group A in the immediate testing condition is reliably greater than zero ($t = 2{\cdot}43$; D.F. 15; $P < 0{\cdot}05$). The mean *s*-score for this group does not change over the 2-day interval ($t = 0{\cdot}76$; D.F. 15; N.S.), and the mean *s*-score of group B is not reliably different from that of group A in the immediate condition ($t = 1{\cdot}14$; D.F. 30; N.S.). Again, there is no evidence of increase of simplification over time; in fact, there is a slight decrease for group A, which is consistent with the findings of Expt. I.

The scores for groups C and D were analysed at greater length than those of groups A and B, because the degrees of freedom in the subject's response pattern are greater in the recall testing than in the recognition testing. N, D, N', and D' can all vary independently, in the sense that any dependence that may exist between them is due to the psychological factor of how well the stories are understood, rather than to any logical necessity. The subject himself is even free to be logically inconsistent. Thus, four separate scores may be analysed.

In order to score the recall answers, it was necessary to adopt a number of criteria for resolving ambiguities; these criteria were as follows.

(i) Not to count redundant assertions. For example, if the subject stated 'No ranchers are farmers', his further statement that 'No farmers are ranchers' was not counted.

(ii) Not to count tautologous assertions. For example, the statement 'all urban Communist representatives are from the urban area' was not counted.

(iii) Not to count assertions about non-existent sets. For example, the statement 'all manufacturers were members of the pro-canal association' was not counted.

(iv) To count assertions about dual sets as two separate assertions. For example, the assertion 'all Communist representatives voted for farm subsidies' was counted as the two statements 'all urban Communist representatives voted for farm subsidies' and 'all rural Communist representatives voted for farm subsidies'.

(v) Not to count an assertion of the form 'not all but some X are Y' when the subject in addition made the assertion that 'all Y are X'. For example, if the subject made the statement that 'all farmers voted for the canal', his further statement that 'not all but some people who voted for the canal were farmers' was not counted.

With the exception of (v), these criteria rarely needed to be invoked.

$(N + D)$, the number of correctly recalled relations, is again termed a *c*-score. Table 2*c* gives mean *c*-scores for the two recall groups; there are no reliable differences among them.

$(N' - D')$, the number of overgeneralizations minus the number of pseudo-discriminations, is again termed an *s*-score; in this case $(N' - D')$ is not necessarily equal to $(N - D)$. Table 2*d* gives mean *s*-scores for the two recall groups. An *s*-score of zero is expected on the basis of chance alone, since there are an equal number of disjunctive and nested relations among the sets listed. The mean *s*-score of group C immediately after reading is reliably greater than zero ($t = 4{\cdot}43$; D.F. 15; $P < 0{\cdot}001$). There are, however, no reliable differences among the groups.

The two aspects of the subjects' responses already analysed have been $(N + D)$ and $(N' - D')$. Here, a new score is proposed; it is $(N + D) - (N' + D')$, the number of correct assertions minus the number of incorrect assertions, and is termed an *a*-score (or accuracy score). Table 2*e* gives *a*-scores for the two recall groups. Accuracy is poor; in no case is the mean *a*-score significantly greater than zero. Moreover, there are no reliable differences among the groups.

The final (independent) score analysed is an *s'*-score (or simplicity score); it is equal to $(N - D)$, the number of correctly recalled nested alternatives minus the number of correctly recalled disjunctive alternatives. Mean *s'*-scores for the recall groups are presented in Table 2*f*. First, the *s'*-score to be expected by chance alone is zero, since there are again an equal number of nested and disjunctive alternatives. The mean *s'*-score of group C in the immediate testing condition is reliably greater than zero ($t = 3{\cdot}39$; D.F. 15; $P < 0{\cdot}01$). The other reliable finding is that the *s'*-score of group D is significantly *less* than the *s'* score of group C in the immediate condition ($t = 2{\cdot}11$; D.F. 30; $P < 0{\cdot}05$).

Discussion

The subjects' accuracy in the recall task is poor; the mean a-score immediately after reading is effectively zero, while the mean a-score of the group taking only the delayed testing is negative. This poor performance is compatible with Bartlett's finding that accuracy of recall is 'not the rule' (Bartlett, 1932, p. 93). The present finding is more striking than Bartlett's, however, because it is not concerned with recall of specific verbal units, but rather with recall of the basic relations asserted by the material. Bartlett also presents examples in which the whole meaning of a passage is distorted. (In contrast, the subjects' accuracy on the recognition task is much better—at least reliably greater than chance.)

A subject's simplification is measured by his s-score (number of overgeneralizations minus number of pseudo-discriminations) and, if he is in a recall group, by his s' score (number of correct nested assertions minus number of correct disjunctive assertions). Simplification occurs. The mean s-scores of all groups in both experiments are reliably greater than zero. Thus, distortion tends to yield simplification. Moreover, the mean s'-scores of the successive recall group in Expt. II is reliably greater than zero. Thus, the selective recall of these subjects tends to yield simplification.

But simplification does not increase over time. In the recognition condition in both experiments, the s-scores actually decrease for subjects who are tested both immediately and after a delay. The results in the recall conditions are not so clear-cut, but there is certainly no tendency for an increase in s scores. Nor do s'-scores increase; in fact the s'-scores of group D are significantly *lower* than those of group C immediately after reading.

The previous findings that simplification increased over time may be due to the fact that simplification and forgetting are confounded in the method of reproduction. That is, by simply omitting reference to certain sets in his reproduction of meaningful material, the subject appears to have simplified it.

CONCLUSION

The system for measuring distortion discussed in this paper is not limited to measuring distortion of written material. It has been shown (Dawes, 1964) that logical fallacies as well may be classified as either overgeneralizations or pseudo-discriminations, and that here too, overgeneralization predominates. Finally, the point that overgeneralization increases structure yields a plausible reason for expecting overgeneralization to be a more common distortion than pseudo-discrimination, whether the distortion is of written materials or of other input. Garner (1962) has written: 'the search for structure is inherent in behaviour' (p. 339). He was discussing probabilistic structure, measured with the statistics of information theory. This paper has discussed *semantic* or *logical* structure, measured in terms of set relations; it has been shown that subjects' distortion of meaningful written materials tends to increase the structure in that material.

I should like to thank Dr Melvin Manis and Miss Barbara Gans for their valuable criticisms of this research and this paper. I should also like to express my appreciation of Mrs Mary Trickett's careful clerical assistance.

REFERENCES

BALLARD, B. P. (1913). Obliviscence and reminiscence. *Brit. J. Psychol. Monogr. Suppl.* 1 (no. 2), 1–82.

BARTLETT, F. C. (1932). *Remembering: a Study in Experimental and Social Psychology.* Cambridge University Press.

DAWES, R. M. (1964). Cognitive distortion. *Psychol. Rep.* (Monogr. Suppl. 4–VI4) 14, 443–59.

GARNER, W. R. (1962). *Uncertainty and Structure as Psychological Concepts.* New York: Wiley.

GERGONNE, J. D. (1817). Essai de dialectique rationelle. *Annales des mathematiques pures et appliqués,* 7.

GORDON, K. (1925). Class results with spaced and unspaced memorizing. *J. exp. Psychol.* 8, 337–43.

HENDERSON, E. N. (1903). A study of memory for connected trains of thought. *Psychol. Monogr.* 5, no. 23, 1–92.

WILLIAMS, O. (1926). A study of the phenomenon of reminiscence. *J. exp. Psychol.* 9, 368–87.

(*Manuscript received* 6 *July* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 87–91
Printed in Great Britain

RELATION BETWEEN SEMANTIC SATIATION AND VERBAL CONDITIONING

BY J. P. DAS*
University of California, Los Angeles

The relation between extent of semantic satiation and trials to learn two verbal conditioning tasks was investigated in a sample of thirty-eight college students. Semantic satiation was measured by polarity-difference scores between rating the words before and after a standard satiation treatment. Trials to reach criterion of learning were combined for the two verbal conditioning tests. The correlation between the satiation and verbal conditioning scores was $r = -0{\cdot}484$; the mean satiation scores of the subjects, dichotomized at the median as fast and slow learners, were also significantly different. These results are consistent with the prediction that facility in verbal conditioning, like susceptibility to hypnosis, depends on resistance to semantic satiation. A subsidiary finding was a significant correlation, $r = 0{\cdot}42$, between the two verbal conditioning test scores.

The prediction that proficiency in verbal conditioning and resistance to semantic satiation should be positively related was formulated on the basis of results of a recent study (Das, 1964*a*). Among other things, this experiment investigated the relation between hypnotic susceptibility and semantic satiation, which were found to be strongly correlated. It implied that subjects highly susceptible to hypnosis tend to have low satiation scores. This, together with the finding that hypnotic susceptibility is related to scores on eye-lid conditioning (Das, 1958) and to verbal conditioning (Weiss, Ullman & Krasner, 1960), leads to the anticipation that verbal conditioning, like hypnosis, would correlate with satiation. It is possible that in hypnosis, as in verbal conditioning, a few verbal stimuli are repeated over and over again; anyone for whom these repetitive stimuli cease to act as meaningful signals cannot have high scores either in hypnosis or in verbal conditioning. Thus proficiency in verbal conditioning will depend on how far an individual can attend to the stimuli without becoming satiated. Although the assumption of a relation between verbal conditioning and hypnosis is not necessary for the present study, it nevertheless provides the rationale for the experiment.

METHOD

Tasks and procedure. Two verbal conditioning tasks and a test of semantic satiation were administered to all subjects.

One of the verbal conditioning tasks was Taffel's (1955) 'Pronoun test', where on each trial the subject was asked to choose one of four pronouns and the only verb in the past tense that was available, and to make up a sentence. The test had been used in the past (Das & Mitra, 1962) with one modification, which was that subjects were asked to make up a two-word sentence using just one of the pronouns and the verb, whereas in Taffel's study subjects were not restricted to the use of only the two words. Latencies of response will be relatively stable with the Das & Mitra procedure and subjects learn faster than under Taffel's original procedure. Subjects were asked in the present study to make up a two-word sentence, and were given sixty trials in the Pronoun test. One-half of the subjects were reinforced for starting a sentence with 'I' or 'We', and the other half for starting a sentence with 'He' or 'They'. A subject found before him a circular card with the four pronouns written round the circle, and he was asked to choose one of them to make up a sentence as soon as he heard the verb from the magnetic tape. Sixty verbs in the past tense were recorded on tape, and were presented to the subject one at a time. Ten seconds after the subject had given his response, the next verb was presented. Every response was reinforced with 'right' or 'wrong', depending upon whether the subject had picked up one of the correct or incorrect pronouns according to the experimenter's scheme.

---

* Now at Utkal University, Bhubaneswar-4, India.

The other verbal conditioning task, the adjective test, was a modification of a previously used test (Das, 1961), and consisted of anticipating whether the word 'good' or 'bad' will follow when one of the two stimulus words, 'Tes' and 'Tig', were presented. The latter were nonsense syllables of comparable low association value taken from Underwood & Schulz (1960). Both nonsense syllables were recorded forty times on magnetic tape, in a predetermined random sequence. As soon as any one of these stimulus words was heard, the subject responded by anticipating either 'good' or 'bad'; he then heard 'good' or 'bad' from the tape, and the subject announced whether his guess was right or wrong. A fixed interval of 10 sec was allowed between each of these operations. One half of the subjects were arbitrarily assigned to the 'Tes-good, Tig-bad' group; the other half to the 'Tes-bad, Tig-good' group.

In the Pronoun test, the subject's choice of pronoun and the latency of his choice was recorded on every trial. Similarly, in the Adjective test both the subject's responses (good or bad) and their latency were recorded. The two tests were administered in a counter-balanced order, but all subjects were tested for semantic satiation after the two tests of verbal conditioning. This was done to avoid any residual effect of satiation on verbal conditioning.

The test of semantic satiation was essentially the same as that used in a previous study (Das, 1964*a*). The five stimulus words chosen for satiation treatment were 'child', 'me', 'rich', 'family' and 'truth'. These were given to subjects for rating on six semantic-differentials two each from the dimensions of 'evaluation', 'potency' and 'activity'). Subjects received the five words in one of five predetermined random orders. As before, each word was presented for repetition at 2–3 sec for a period of 40 sec and was rated again on the same scales immediately after its repetition. The order in which the words were rated before repetition was followed for repetition and the rating that followed it. An interval of 30 sec was allowed between the end of rating one word and repetition of the next word.

A new phase was added to the test after the final rating of all five words. This consisted of obtaining associations for the satiated words. Each of the satiated words was presented with the instruction that the subject should give as many associations to the word as he could until he was stopped at the end of a minute. Latency for the first association to each was recorded. Introspective reports were taken at the end of the verbal conditioning tests, as well as at the end of the satiation test.

The main stages in the procedure may be summarized at this point. Two verbal conditioning tasks, separated from each other by 5 min, were administered to subjects; half of them took the Pronoun test first, the other half took the Adjective test first. Semantic satiation was induced and measured in a standard procedure after verbal conditioning. Then the satiated words were presented individually in order to elicit free association response words for 1 min.

Subjects. Forty male and female volunteers participated in the experiment as a class requirement in introductory psychology.

Apparatus. The subject, and the experimenter with the control equipment were in adjacent rooms fitted with a one-way vision window. The equipment was used to present the verbal stimuli, pre-recorded on magnetic tape, to the subject and to measure the subject's response latencies. A tape recorder, Hunter's 'Klockounter and interval timer', a voice key and microphones were used. The recorded latency of response was the interval begining when a stimulus from the tape started off the clock, and ending when a subject's oral response stopped the clock. A fixed period of 10 sec intervened between a subject's response and the next stimulus from the tape. Communication with subjects was maintained through an intercom system.

RESULTS AND DISCUSSION

The relation between semantic satiation and verbal conditioning is discussed first. Some subsidiary findings concerning the tests of verbal conditioning and of satiation are then discussed separately.

Semantic satiation and verbal conditioning

The learning score on each verbal conditioning test was the number of trials to reach a criterion of eight consecutively correct responses. A subject was classed as a non-learner if the criterion was not reached by 60 trials in the Pronoun test, and by 40 trials in the Adjective test. A combined learning score for each subject was also obtained by adding his scores on the two tests. The latter scores for thirty-eight subjects* were divided at the median, separating them into nineteen fast and nineteen slow learners (FL and SL). Comparison of the mean satiation scores (polarity difference) of the two groups revealed that the FL group had a significantly higher score (mean $= +1\cdot31$, S.D. $= 5\cdot60$) than the SL group (mean $= -5\cdot42$, S.D. $= 4\cdot50$); $t = 4\cdot08$, D.F. $= 18$ $P < 0\cdot001$. Following the prevalent practice of describing positive and negative satiation scores (cf. Lambert & Jakobovits, 1960; Yelen & Schulz, 1963), the FL group demonstrated semantic *generation*, and the SL group semantic *satiation*, However, such a dichotomy is not very meaningful, since each mean was derived from a set of positive and negative polarity-difference scores. The means imply that the FL group displayed stronger resistance to satiation than the SL group.

Essentially the same relation is expressed by the product–moment correlation of $r = -0\cdot484$ between the combined learning (error) scores and the scores on semantic satiation of all thirty-eight subjects. High error scores were associated with negative satiation scores. This is consistent with the relation observed between hypnotic susceptibility and resistance to satiation ($r = 0\cdot667$; Das, 1964*a*). Thus the results confirm the prediction that efficiency in verbal conditioning may depend partially on the magnitude of resistance to satiation. The large error variance in the combined learning scores perhaps limits the chances of obtaining a strong relation.

A second measure of semantic satiation was associative response time (RT) to the satiated word presented as a stimulus. The score was the RT for the first response word on presentation of the stimulus word. The total RT score for a subject was simply the sum of the RT for all five satiated words. Following the same dichotomy of FL and SL, the mean difference in RT for the two groups gave $t = 1\cdot70$ (D.F. $= 18$; $0\cdot20 > P > 0\cdot10$). The means for the FL group (11·89) was shorter than that for SL group (14·03), as would have been expected from the correlation between learning and resistance to satiation; but its failure to reach statistical significance again suggests, as was found before (Das & Cook, 1964), that associative RT to a satiated word is an inadequate measure of semantic satiation. In fact, the two measures of satiation were found to have little relation with each other: the product–moment correlation between polarity-difference scores for the thirty-eight subjects and total RT was 0·07. The two measures probably represent, respectively, connotative and associative satiation (Amster, 1964), and the near-zero correlation is suggestive of a lack of close correspondence between the two types of satiation.

---

* Records of two subjects were eliminated. One subject did not co-operate during testing, giving random responses and refusing to give free associations to the five satiated words. The other subject was found to have incomplete latency records.

Verbal conditioning scores

An interesting finding was a positive correlation, $r = 0{\cdot}42$ ($P < 0{\cdot}01$) between the Pronoun and Adjective test scores. An earlier study suggested that verbal conditioning tests, like those of classical conditioning, might not correlate with each other (Das, 1961); in this study one of the tasks was very similar to the Adjective test, while the other was a test of pattern discrimination with verbal cues. Apparently, those two tests were less similar to each other than the present tests. This may prompt one to ask if the correlation obtained here was merely due to a positive transfer from one test to the other irrespective of which test was given first. In other words, it may be that experience with the first test facilitated performance in the second. Comparison of the means of the Pronoun test given first versus second showed no gain for the latter ($t = 1{\cdot}29$; D.F. $= 36$; $P > 0{\cdot}20$). For the Adjective test, the corresponding mean difference was almost significant: $t = 2{\cdot}0 < 2{\cdot}027$ at $P = 0{\cdot}05$. It may be concluded that although some positive transfer might have contributed towards the correlation between the two tests, the further assumption that the two tests are independent measures of some common ability in verbal conditioning would be necessary to account fully for the substantial relation obtained here between the two tests.

Satiation scores

The mean polarity-difference score for the sample was $-1{\cdot}98$, summed over five words and six scales. This compared favourably with scores from other satiation studies on monolingual samples. Kanungo & Lambert (1963) had a mean of $-1{\cdot}92$ for a Canadian sample. However, in the classical study of Lambert & Jakobovits (1960), the mean was considerably higher ($-9{\cdot}51$). Differences in the choice and number of semantic-differential scales used, and minor differences in procedure and sample may make these results not strictly comparable.

Recently, the adequacy of polarity-difference scores as a measure of satiation has been questioned (Schulz, Weaver & Radtke, 1965). These investigators found no evidence of relative loss of meaning for a set of words when repeated by an experimental group, but not repeated by a control group in a 'post-test only' design. Their results go against the theory that satiation is a type of cognitive inhibition. This is acceptable without any prejudice to the use of polarity-difference scores which are derived from the difference between pre- and post-repetition ratings. These scores, as index of satiation, have given rise to stable relations with verbal conditioning, hypnosis, and vigilance.

Separate computations of the latency of first association for each of the five words showed that the mean latencies in seconds for 'truth' was the longest (3·01) followed by 'me' (2·88), 'child' (2·55), 'family' (2·30), and 'rich' (2·23). The most frequent first association for these words were: truth–good ($n = 5$), me–I ($n = 5$), child–small (7), family–mother (8), and rich–money (16). The word 'rich' appears to be the most familiar, and has the shortest associative latency; it also happens to have the highest negative polarity-difference score. The mean polarity-difference scores for the individual words (summed over the six semantic differentials are: truth, $+0{\cdot}03$; me, $+0{\cdot}34$; child, $-0{\cdot}95$; family, $-0{\cdot}03$; rich, $-1{\cdot}37$. Since the observations are so few, no correlational statistics can be used. The data merely serve to suggest that familiar words are easily satiated, a finding reported in previous studies on satiation (Das, 1964*b*).

The introspective reports following verbal conditioning were analysed with a view to detect learning without awareness. All subjects who learned a task could verbalize the simple principle of the test. However, there were two subjects who verbalized the principle correctly, but who did not reach a criterion of learning. The typical answer given by them indicated that they did not care for the experimenter's verbal reinforcements.

The introspective reports following satiation were not particularly informative except in a few cases where subjects wanted to explain why pre-repetition rating would be likely to change following repetition. A new explanation given by these subjects was that they attended to the *meaning* of the word on the first rating, but to the *sound* of the word on the second rating. Experiments deliberately manipulating this variable may prove to be important in the theoretical interpretation of semantic satiation.

The author wishes to thank Prof. C. Y. Nakamura for permitting the use of his equipment and laboratory, and to the Department of Psychology, UCLA, for financial support.

REFERENCES

AMSTER, H. (1964). Semantic satiation and generation: learning? adaptation? *Psychol. Bull.* **62,** 273–86.

DAS, J. P. (1958). Conditioning and hypnosis. *J. exp. Psychol.* **56,** 110–3.

DAS, J. P. (1961). Some correlates of verbal conditioning. *Psychological Studies* **6,** 30–35.

DAS, J. P. (1964*a*). Hypnosis, verbal satiation, vigilance and personality factors: a correlational study. *J. abnorm. soc. Psychol.* **68,** 72–8.

DAS, J. P. (1964*b*). Semantic satiation: an experiment and a theory. *Peabody Papers in Human Development,* **2,** no. 5. George Peabody College for Teachers. Nashville.

DAS, J. P. & COOK, C. H. (1964). On the development of an associative response time (RT) measure for semantic satiation. *Psychol. Rep.* **15,** 52.

DAS, J. P. & MITRA, A. K. (1962). Relative effectiveness of electric shock and praise and reproof in verbal conditioning. *J. gen. Psychol.* **67,** 141–6.

KANUNGO R. & LAMBERT, W. E. (1963). Paired-associate learning as a function of stimulus and response satiation. *Brit. J. Psychol.* **54,** 135–44.

LAMBERT, W. E. & JAKOBOVITS, L. A. (1960). Verbal satiation and changes in the intensity of meaning. *J. exp. Psychol.* **60,** 376–83.

SCHULZ, R. W., WEAVER, G. E. & RADTKE, R. C. (1965). Verbal satiation? *Psychonom. Sci.* **2,** 43–5.

TAFFEL, C. (1955). Anxiety and the conditioning of verbal behavior. *J. abnorm. and soc. Psychol.* **51,** 406–501.

UNDERWOOD, B. J. & SCHULZ, R. W. (1960). *Meaningfulness and Verbal Learning.* Chicago: Lippincott.

WEISS, R. L., ULLMAN, L. P. & KRASNER, L. (1960). On the relationship between hypnotizability and response to verbal operant conditioning. *Psychol. Rep.* **6,** 59–60.

YELEN, D. R. & SCHULZ, R. W. (1963). Verbal satiation? *J. verb. Learn. verb. Behav.* **1,** 372–7.

(*Manuscript received* 19 *July* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 93–97
Printed in Great Britain

THE EFFECT OF SHORT PERIODS OF FOOD-DEPRIVATION ON HUMAN PERFORMANCE

By R. A. KENNEDY* and A. G. KEENE
University of Melbourne

This study reports the effect of periods of food-deprivation varying from 0 to 10 hr on human performance in two simple tasks. The argument is advanced that an interpretation of the results in terms of traditional Hullian multiplicative drive theory is unsatisfactory. A possible alternative to this approach is considered.

There are comparatively few studies of the effects of short periods of food deprivation on the performance of human subjects. In the main, these have directed attention towards selective effects in perception, and systematic changes in word-association tasks (Brown, 1961). However, information on the effects of food deprivation in a variety of situations, including more complex learning tasks, is of some importance for traditional Hullian theory (Hull, 1943, 1951). In particular, this theory would assume food deprivation to contribute directly to strength of drive (D), which, in turn, is related to performance in a monotonic fashion. Thus, if the effects of inanition may be ignored over short periods of time, performance would be expected to improve steadily as the level of deprivation increased from, say, 0 to 10 hr.

There is now a considerable body of evidence for a curvilinear relation between various measures of 'arousal' and performance (Malmo, 1958). This arousal theory, which is to a degree physiologically committed, suggests that an optimal level of cortical activity is necessary for the performance of a task. The Hullian theorist with regard to the mechanism of the hunger drive in human subjects would presumably speak in terms of some substance present in the blood which bathes the neural structures constituting ${}_SH_R$, and thereby 'sensitizes' them, increasing the probability of their responding. Although the argument seems somewhat tenuous it could be suggested for the arousal theorist that stimuli characteristic of increasing deprivation would serve to raise the level of cortical activity. Both theoretical systems remain viable in spite of their physiological encumbrances. Malmo (1958) equated 'arousal' and 'drive', suggesting a relation between drive and performance distinct from that of the Hullian theorist. Differential prediction is made difficult by the fact that obtained curvilinearity in the relation of drive to performance may be derived in the Hullian case by the postulated positive relation between inanition (ϵ) and deprivation; and also by the fact that D may be expected to enhance not only 'correct' responses, but also competing ,'incorrect', response tendencies.

This paper discusses the performance of human subjects in two situations: a relatively simple checking test (*Minnesota Clerical Test*, Andrew & Paterson, 1946), and a more complex paired-associate verbal learning task.

For the checking test, the arousal theory would predict an 'inverted U' relation, the exact position of the optimal level being a function of the difficulty of the task, and open to determination. In this case it is difficult to see how a multiplicative theory of drive can predict anything other than an improvement in performance with increasing periods of deprivation, up to the time when the effects of inanition are more than negligible. Evidence on the inanition component in Hullian theory is rather scant: Hull (1951) quoted a study made by Yamaguchi (1951) which showed a decrement in the performance of rats (presumably due to inanition) after 60 hr of food-deprivation. An experiment by Horenstein (1951) examining in detail the effects of up to 24 hr of deprivation in rats showed no evidence of a decline in performance, though the most rapid increase was found after only 2 hr. While it is easy to appreciate that long periods without food will affect the performance of a demanding physical task, it is harder to see why this should occur with relatively short periods of deprivation, and with simple laboratory tasks calling for no physical effort. In any case no clear statement on the mechanism of inanition in human subjects is available.

---

* Now at Department of Psychology, Queen's College, Dundee.

In the paired-associate learning situation prediction is more complicated. The arousal position would again predict a curvilinear relation between drive and performance, with an optimal point somewhat below that for the less complex checking task (Bindra, 1959). Spence (1956), using a neo-Hullian model, has argued that, for paired-associate tasks where the response words are anything but common high-frequency associates of the stimulus elements, increases in drive may lead to decrements in performance. This follows from his argument that increasing drive raises the strength of all response tendencies, which effectively increases the number of incorrect anticipations made. The subject in this situation, making more errors, learns more slowly. Since the word pairs used in the present study contained no obvious semantic or syntactic dependencies, Spence's position would lead to the prediction that with longer periods of deprivation (higher drive), more errors would be made, and more trials taken to reach a criterion. However, the drive variable and the word lists are unlike any that Spence has used, and on a number of occasions he has forcibly made the point that derivations from his theory must satisfy certain 'boundary conditions' with regard to experimentation. This study is not intended to be a test of his position; the extrapolations from his work are those which would be most damaging to the interpretation of the present results. A prediction of curvilinearity cannot be derived from his theory with a word list of the form used here.

METHOD

Subjects

These were thirty female students in a first-year psychology course; their ages ranged from 17 to 21 yr. All were unfamiliar with paired-associate learning. No subject was under medical supervision or was dieting at the time of the experiment.

Tasks

The checking task used was part of the *Minnesota Clerical Test* (Andrew & Paterson, 1946).

The word list was a six item list composed of two-syllable words selected from the lists of medium difficulty presented by Greenburg & Underwood (1950). It was recorded fifteen times on magnetic tape. Ten different random orders were used, five of these being repeated. The stimulus-response sequence took the form: stimulus word, followed by a pause of 3·5 sec; the stimulus-response pair; a pause of 2 sec; and the next stimulus item. An inter-trial interval of 5 sec was used, the subject responding during the 3·5 sec pause. This response, together with the original list, was recorded on a second magnetic tape and used in a study of response latencies not reported here. The experimenter also made a written record of the responses.

Procedure

The thirty subjects were randomly assigned to three groups of ten. All the groups were given details of the times at which they were to eat on the day of the experiment, and the times to report. One subject from each group was tested each weekday for 10 days. Group 1 (10 hr deprivation) ate at 8·00–8.15 a.m. and reported at 6.15 p.m.; group 2 (5 hr deprivation) ate at 1.00–1.30 p.m. and reported at 6.30 p.m.; group 3 (no deprivation) ate at 6.15–6.30 p.m. and reported at 6.45 p.m. Subjects were allowed to eat one of a choice of three meals, each of which had a calorific content of about 800. The meals were made up mainly of bread, butter and milk. They were asked to drink only water between the time of their last meal and the experiment.

In each testing session the subject first completed the checking test, and then performed the paired-associate learning task. Standard instructions for the *Minnesota Clerical Test* were given, the score recorded being the number of items attempted in 4 min (8 min are allowed if the norms of this test are to be used). Normal instructions for paired-associate learning were given, including a one-trial practice session on a three-item list. Three subjects (two in group 2, and one in group 1) were lost either through apparatus failure or through failure to observe instructions. Three further subjects were dropped at random in the analysis of the learning data in order to provide three equally sized groups.

RESULTS

On the *Minnesota Checking Test* the mean score of the 0 hr group was 68·1; of the 5 hr group, 78·8; and of the 10 hr group, 61·4. There is an apparent 'inverted *U*' relation between performance and hours of deprivation. As the distributions of scores were markedly non-normal, and the groups were small, these data were analysed using non-parametric tests. A Kruskal–Wallis analysis of variance over the three groups gave $H = 6{\cdot}25$ ($P < 0{\cdot}05$). The differences between pairs of means were tested using the Mann–Whitney U statistic. The 0 hr and 5 hr groups, and the 5 hr and 10 hr groups, were significantly different ($z = 1{\cdot}65$, $P < 0{\cdot}05$; $z = 2{\cdot}34$, $P < 0{\cdot}01$). The 0 hr and 10 hr groups were not significantly different.

Table 1 shows the mean and S.D. of the total numbers of correct responses made by the three groups on each trial of the learning task. The apparent differences between the groups were examined in an analysis of variance of the combined scores on the first two and the last two trials (cf. Table 2). Since the list was very easy to learn, any differences could be expected to occur only during the early trials; analyses of variance on the first two and the last two trials separately, together with Newman–Keuls analyses of differences between means, showed no differences between groups

Table 1. *Means and* S.D. *of correct responses for the first two and the last two trials of a paired-associate learning task by groups of human subjects deprived of food for* 0, 5 *and* 10 *hr* ($n = 8$/*group*)

		Trial			
Deprivation group		1	2	13	14
0 hr	Mean	2·50	3·25	6·00	6·00
	S.D.	0·93	1·28	0·00	0·00
5 hr	Mean	0·75	1·375	5·75	6·00
	S.D.	0·46	1·19	0·46	0·00
10 hr	Mean	1·375	2·625	5·625	6·00
	S.D.	1·19	1·41	0·52	0·00

on the last two trials. On trial 1, the 0 hr group performed significantly better than the 5 hr and 10 hr groups ($P < 0·05$). The 5 hr and 10 hr groups were not significantly different. On trial 2, only the 0 hr and 5 hr groups differed significantly ($P < 0·05$).

An analysis of variance of the numbers of errors made by the three groups gave clearly non-significant results ($F = 1·05$; D.F. = 2 and 21; $P > 0·20$).

Table 2. *Analysis of variance of the combined correct responses for the first two and the last two trials of the paired-associate learning task*

Source	D.F.	S.S.	M.S.	F	P
Between subjects (S)	23	45·6	—	—	—
Deprivation (D)	2	15·0	7·50	5·14	$< 0·05$
Subjects within groups	21	30·6	1·46	—	—
Within subjects	72	426·0	—	—	—
Trials (T)	3	377·9	125·96	244·93	$< 0·001$
$D \times T$	6	12·7	2·12	3·76	$< 0·01$
$T \times S$ within groups	63	35·4	0·56	—	—
Total	95	471·6	—	—	—

DISCUSSION

The findings of this study are difficult to interpret in terms of a multiplicative drive theory. The obtained curvilinearity of the relation between deprivation time and performance cannot be accounted for in terms of growing inanition with increasing food-deprivation. This follows from the fact that the relative performance of the three groups differs on the two tasks. The 0 hr group performed best in the learning situation, and the 5 hr group in the checking task.

Using reasoning similar to that of Spence one might attempt to account for the relatively poorer performance of the 5 hr group in learning the paired-associate list, but only by suggesting that they would make more errors. However, differences in the number of errors made by the groups are not significant. According to the Hullian formulation, the combined effects of drive and inanition are a function of the amount of food-deprivation, and would therefore influence performance on both tasks equally. It is not possible for the optimal period of deprivation to change for the two tasks. A possible weakness of the foregoing argument is that measures of errors in paired associate learning are influenced greatly by the instructions given prior to learning. This certainly limits the usefulness of comparisons between experiments; but since identical instructions were given to all groups in the present study, and, further, *some* errors were elicited from all groups, the comparisons between groups are considered to be legitimate.

It would remain open to a Hullian theorist to postulate the existence of responses, attached to drive stimuli (S_D), which are incompatible with the performance of one task and not with the other; but there is no evidence that such responses occur, or what their nature would be, and this *ad hoc* explanation would still have to account for the relative changes in the effects of these responses with increasing deprivation for the two tasks.

A more parsimonious interpretation would appear to lie in a postulated curvilinear relation between short periods of food-deprivation and performance; the optimal level being lower for the more complex task. In this context, it must be recalled that 0 hr of food-deprivation cannot be equated with zero level of arousal. Stimuli characteristic of satiation would be expected to contribute to the level of cortical activity, as would various other sources of stimulation.

The findings of this study demonstrate measurable changes in performance on simple tasks with relatively short periods of food-deprivation. In the light of them, Malmo's comments (1958) on Franks' work (1957) appear reasonable. He noted that with two groups of subjects, deprived for either 0 or 18 hr of food, drink, and tobacco, no differences in the rate of acquisition of a conditioned response would be expected, since the optimal point of arousal between these two points had possibly been missed. This does not, of course, alter the original interpretation that Franks placed on his results illustrating the inapplicability of the Hullian model in situations of this sort. However, his contention that the absence of differences between groups could be taken as evidence for the introversion–extraversion theory of conditioning gains no support from the present results.

The present results may also be taken to lend further support to the Yerkes–Dodson Law, (Yerkes & Dodson, 1908). This well-known statement of the relation between task difficulty and motivation has been given consistent, if sporadic, support over the past fifty years, recently for instance by Broadhurst (1959). Inasmuch as the present data could be said to stem solely from the manipulation of appetitive drives the law is given support in a hitherto neglected area (Cofer & Appley, 1964, p. 523).

References

Andrew, D. M. & Paterson, D. G. (1946). *Minnesota Vocational Test for Clerical Workers.* New York: Pychological Corporation.

Bindra, D. (1959). *Motivation: A Systematic Reinterpretation.* New York: Ronald Press.

Broadhurst, P. L. (1959). The interaction of task difficulty and motivation: The Yerkes–Dodson Law revived. *Acta Psychol.* **16**, 321–38.

Brown, J. S. (1961). *The Motivation of Behaviour.* New York: McGraw-Hill.

Cofer, C. N. & Appley, M. H. (1964). *Motivation: Theory and Research.* New York: Wiley.

Franks, C. M. (1957). Effect of food, drink and tobacco deprivation on the conditioning of the eyeblink response. *J. exp. Psychol.* **53**, 117–20.

Greenburg, R. & Underwood, B. J. (1950). Retention as a function of stage of practice. *J. exp. Psychol.* **40**, 452–7.

Hull, C. L. (1943). *Principles of Behaviour.* New York: Appleton-Century-Crofts.

Hull, C. L. (1951). *Essentials of Behaviour.* New Haven: Yale University Press.

Horenstein, B. R. (1951). Performance of conditioned responses as a function of strength of hunger drive. *J. comp. physiol. Psychol.* **44**, 210–24.

Malmo, R. B. (1958). Measurement of drive: An unsolved problem. In *Nebraska Symposium on Motivation,* ed. M. R. Jones. Lincoln: University of Nebraska Press.

Spence, K. W. (1956). *Behaviour Theory and Conditioning.* New Haven: Yale University Press.

Yamaguchi, H. G. (1951). Drive (*D*) as a function of hours of hunger (*h*). *J. exp. Psychol.* **42**, 108–17.

Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *J. comp. neurol. Psychol.* **18**, 459–82.

(*Manuscript received* 22 *February* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 99–106
Printed in Great Britain

THE EFFECTS OF DISTRACTION ON PURSUIT ROTOR LEARNING, PERFORMANCE AND REMINISCENCE

BY HANS J. EYSENCK AND WARREN THOMPSON*
Institute of Psychiatry, University of London

Five groups of thirty subjects were equated for performance on the pursuit rotor, and were then given massed practice under conditions of no distraction, a little, medium or considerable distraction, as well as a control distracting condition. It was found that performance declined proportionally to the amount of distraction given and that the effect of distraction was on performance only, and not on learning. During a subsequent rest pause half the subjects were given a distracting task, the other half were simply rested; performance after this rest period failed to show any effect of the distracting task on consolidation processes theoretically taking place during the rest period.

The measurement of distraction effects, i.e. the investigation of the effects of a distracting task (P_d) on performance on another task (P_a), was introduced into psychology in the 1880's, and reviews are available by Geissler (1909), Pauli (1930), and Kreezer, Hill & Manning (1954). Most of this work arose from an interest in certain problems of attention. As this concept began to be disregarded by psychological writers in the last thirty years or so, work on distraction began to decline, although many of the original findings, in spite of being based on small samples, inadequate experimental data and poor statistical treatment, are nevertheless capable of verification (Sterky & Eysenck, 1965).

The experiment described in this paper makes use of distraction in order to attack certain theoretical problems in learning theory which are rather far removed from those which motivated the earlier experimenters, and is in line rather with more recent work, such as that of Briggs, Fitts & Bahrick (1957). In particular we were concerned with two questions which have not received an adequate answer in modern learning theory. The first of these relates to the well-documented fact that performance on P_a declines when P_d is simultaneously performed, in rough proportion to the amount of attention required for the execution of P_d (Sterky & Eysenck, 1965); is it only performance which is interfered with, or is there also an interference with learning? Hullian theory, and indeed any form of reinforcement theory as applied to the learning of skills, would suggest the latter. Let us assume that over a period of 5 min two groups performed on the pursuit rotor, one without P_d, the other with P_d, and let us assume further that performance of the P_d group is roughly at a level one third that of the group without P_d. This means that the P_d group will only receive one third of the total reinforcement given to the other group, reinforcement being here taken as meaning 'knowledge of successful performance'. If learning is dependent in any sense on reinforcement, then the P_d group should learn less well than the other group. Equally, if learning is mediated in whole or in part by the successful performance of a given task, then again the no-P_d group should show considerably more learning than the P_d group. Even using a rather old-fashioned type of nomenclature one might think that the more attention an individual can devote to a task which he is attempting to learn, the more successful would be his attempt, so that the no-P_d group, able to concentrate completely on pursuit rotor learning, should learn much better than the P_d group whose concentration is demonstrably affected by P_d.

---

* W. Thompson is now at Vanderbilt University, Nashville, Tenn., U.S.A.

The second problem to be dealt with relates to the events which take place during a rest pause following a period of massed practice. Eysenck (1965) has argued that in pursuit rotor learning a process of consolidation takes place during this rest pause which enables performance to reach a higher level after the rest pause; this hypothesis is an alternative to the Hull–Kimble theory of dissipation of reactive inhibition. Several writers have shown that whatever processes may go on during the rest pause are relatively immune from various types of activities, ranging from talking to the experimenter to rating jokes and pictures, or looking at magazines. It has also been shown, however, that other types of activity, such as carrying out pursuit rotor practice in a mirror has the effect of interfering with consolidation if the activity is carried out during the first few minutes after massed practice (Eysenck, 1965). It is not known whether these different effects are due to the more or less attention-demanding qualities of P_d, or whether the differences are due to the considerable degree of similarity existing between pursuit rotor performance and pursuit rotor performance in a mirror (the term 'distracting task' in this connexion refers to its possible distracting character as far as the process of consolidation is concerned). It seemed likely that an experiment of this kind using a P_d making considerable demands upon attention but nevertheless quite different in nature to pursuit rotor learning would throw light on this problem.

METHOD

Plan

One hundred and fifty male subjects (industrial apprentices), ranging in age from 16 to 18 yr, were divided into five groups of thirty each. Subjects were not naïve with respect to pursuit rotor learning but had taken part in an experiment 12 months previously. Subjects were assigned to groups on the basis of their average level of performance on the pursuit rotor task during an initial practice period extending over 2 min.

Tasks

The experiment involved two tasks, the primary task (P_a) and the distracting task (P_d). The primary task was pursuit rotor learning, in which each subject attempted to keep the tip of a stylus in contact with a 'target' on a rotating turntable. For each of the five groups of subjects, the primary task itself was invariant. Differences between groups involved variations of the distracting task performed simultaneously with the primary task. Pursuit rotor performance was always massed and integrated over 10 sec periods. Details of apparatus, instructions and scoring have been given elsewhere (Eysenck, 1964).

The distracting task consisted of pressing either the right or the left foot pedal in response to one of two distinctive signals, a high- or low-pitched tone. During the experiment, the subject was seated in a chair with each of his feet resting on a pedal. Tones were transmitted to the subject through earphones, and he was instructed to respond to the high-pitched tone by pressing the right foot pedal and to the low-pitched tone by pressing the left foot pedal.

Three levels of difficulty of the distracting task were used, these levels depending upon the frequency with which tones were transmitted to the subject for his response. In the *easy distraction* condition the subject responded to 20 tones/min; the *medium distraction* condition required a response to 47 tones/min; and subjects in the *difficult distraction* condition responded to 72 tones/min. High-pitched tones and low-pitched tones were presented in random order. Further details of this experimental set up, and data regarding its effectiveness, have been given by Sterky & Eysenck (1965).

Practice periods

Initially all five groups were given experience with the primary task and the distracting task. For 2 min, each subject was allowed to practise the primary task. Since the groups were to be equated in terms of the initial ability of the subjects on the pursuit rotor, scores on this initial practice period determined the assignment of the subject to a group, Following this initial practice period, each subject performed the difficult condition of the distracting task for 2 min. After these two practice sessions, subjects were given a 10 min rest period during which they performed an irrelevant task, i.e. ranking a set of polygons according to aesthetic preference.

Experimental period I

During this 5 min period each of the five groups was treated as follows. Group *A* performed the primary task only. Group *B* performed the primary task and the distracting task in the easy condition. Group *C* performed the primary task and the distracting task in the medium condition. Group *D* performed the primary task and the distracting task in the difficult condition. Group *E* (a control group) performed the primary task, but in addition subjects were instructed simply to press either of the foot-pedals once each second; although these subjects heard the tones through the earphones (medium condition), they were instructed to ignore them.

During the last 30 sec of this period, all distracting stimuli were withdrawn, for reasons explained below. After experimental period I, all subjects had a 10 min rest period, during which they again performed the irrelevant task of ranking a different set of polygons.

Experimental period II

The second experimental period consisted of the performance of the primary task only for all five groups for 5 min. After this period each of the groups was divided into two subgroups of fifteen subjects each. One of the subgroups of each pair performed the difficult condition of the distracting task for 5 min; the other subgroup had a 5 min rest period.

Experimental period III

The last experimental period consisted of the performance of the primary task for 2 min by all subjects in each of the five groups.

Instructions

Immediately prior to experimental period I, all subjects were told that during the period the experimenter would say the word, 'Now', which indicated that there were 30 sec remaining in the period. For subjects in group *A*, this signal meant that they were simply to continue the primary task. For subjects in groups *B*, *C* and *D*, this signal meant that they were to discontinue performance of the distracting task and merely perform the primary task. Subjects in group *E* stopped pressing the foot-pedals at the signal and continued the primary task. The use of these instructions enabled the experimenter to obtain a pre-rest measure of a subject's performance in a non-distracted condition.

Scoring

For the primary task, time spent on target was electronically measured during the initial practice period and the three experimental periods; the experimenter recorded these scores for each 10 sec period, each period being termed a trial. In addition, for subjects performing the distracting task, a record of correct, incorrect and omitted responses was kept.

RESULTS

Pursuit rotor performance scores during the initial practice period and experimental periods I and II have been plotted in Fig. 1. It will be seen that all groups perform in almost identical fashion during the 2 min practice period preceding the experiment proper, and analysis of variance confirms a complete absence of significant difference between groups. There is a significant learning effect at the 1 % level, but no significant interaction.

The differential effects of P_d are shown in the first twenty-seven trials of experimental period I, where the groups were graded in precise accordance with the amount of distraction presented. Fig. 2 gives a plot of the performance of groups *A*, *B*, *C* and *D*, showing a linear relation between time on target and mean signal number, which may be taken as an index of distraction. Group *E* performed slightly better than group *C*, although the number of responses on the distracting task produced is much higher; this suggests that a certain amount of distraction is caused by the need to link the

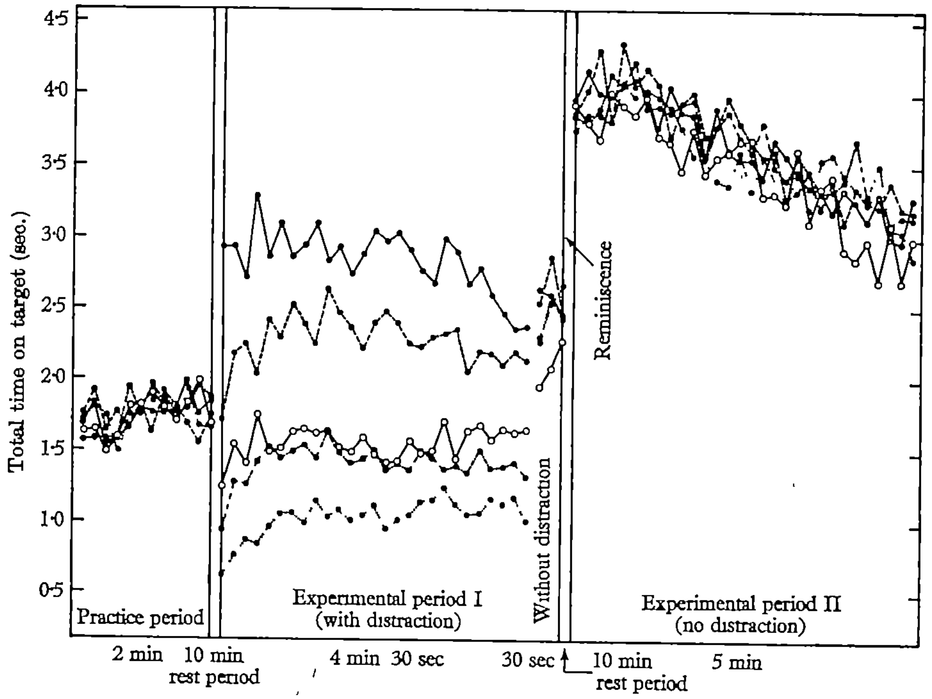


Fig. 1. Pursuit rotor scores for practice period, experimental period I (with distraction), and experimental period II (without distraction). ●——●, Group A (no distraction) ●— — —●, group B (easy distraction) ●— · — ●, group C (medium distraction) ●- - - ●, group D (difficult distraction) ○——○, group E (control group with distracting task).

Table 1. *Analysis of variance of pursuit rotor scores for the twenty-seven trials of experimental period I*

Sources	D.F.	S.S.	M.S.	*F*
Between groups	4	1689·42	422·36	16·714*
Between trials	26	38·61	1·49	3·548*
$G \times I$	104	68·52	0·66	1·571*
People within groups	145	3663·63	25·27	—
Residual	3770	1592·11	0·42	—
Total	4049	7052·29	—	—

* $P < 0{\cdot}01$.

auditory signal to the appropriate foot movement. Analysis of variance (cf. Table 1) gave a highly significant difference between groups, and rather less significant values of *F* (although still beyond the 1 % level) for differences between trials and for the groups by trials interaction. The interaction effect, as can be seen, is due to the fact that group *A* declined in performance throughout the practice period while the other groups (particularly group *D*) improved. The decline of group *A* shows the usually

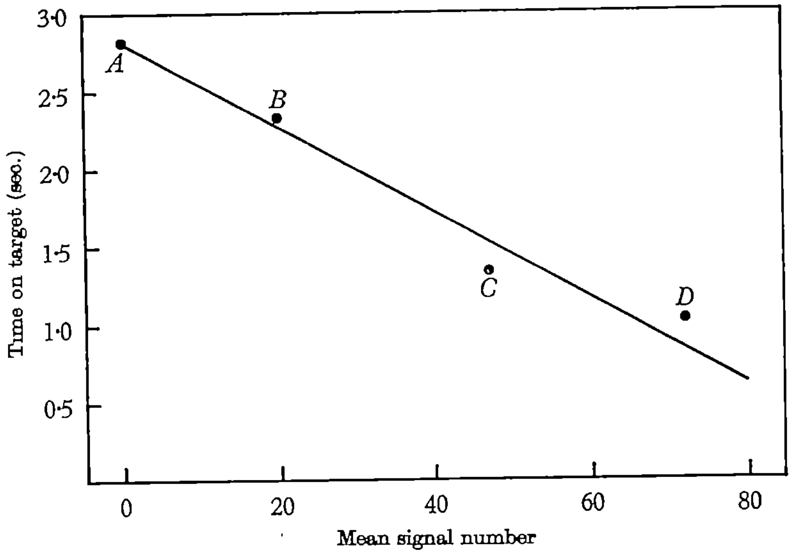


Fig. 2. Differential effect of the severity of distraction (mean signal number per minute) on time on target on pursuit rotor task.

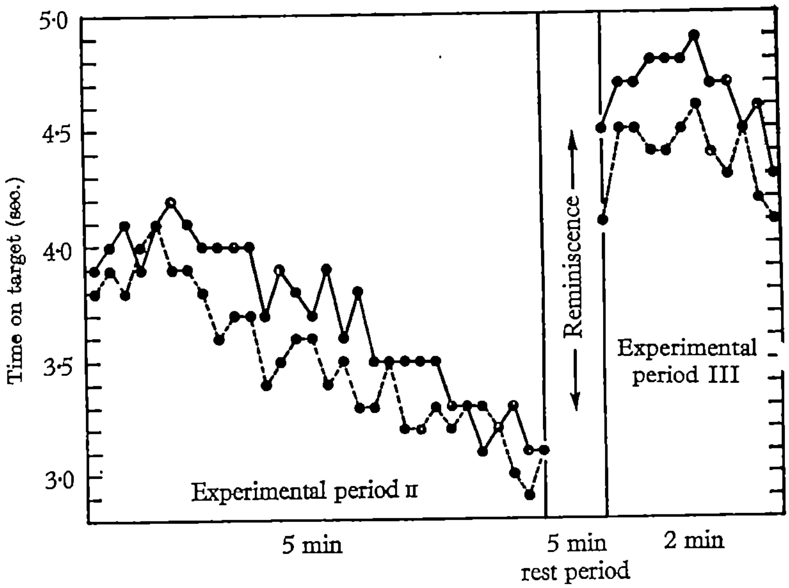


Fig. 3. Performance on pursuit rotor during experimental periods II and III of groups given either distracting task or rest during the 5 min period intervening between experimental periods: – – –, distracting task prior to experimental period III; ——, rest prior to experimental period III.

found phenomenon of post-rest decrement; the unusual increase in scores of the other groups is possibly due to some such effect as learning to disregard the distracting stimulus by making responses more automatic. In order to establish the significance of differences between pairs of groups, Tukey's (1949) test for differences between means was used, and it was found that each mean was significantly different from every other mean at the 1 % level, except that the difference between the means of groups *E* and *C* was significant only at the 5 % level.

All groups worked without P_d for the last three trials of experimental period I. In spite of the difference in pursuit rotor performance during the preceding trials, the five groups are indistinguishable from each other during these 30 sec without P_d ($P > 0{\cdot}05$). Thus it would appear that there has been no interference with learning but that P_d has simply held down performance to a lower level in strict accordance with the amount of P_d imposed.

This finding was brought out equally clearly in experimental period II. All five groups show identical reminiscence and identical performance; none of the differences is statistically significant. The only source of variance giving rise to a significant *F* ratio is that arising from trials; there is a highly significant decline in performance for all groups. Interaction effects are negligible. All data summarized in Fig. 1 agree therefore in demonstrating a failure of P_d to affect learning and the conclusion is suggested that distraction affects performance only.

Fig. 3 shows pursuit rotor performance scores during experimental periods II and III: subjects have been classified into two groups, i.e. the seventy-five subjects (fifteen from each of the five original groups) who were given 5 min rest between experimental periods II and III, and the other seventy-five subjects who performed the medium condition of the distracting task during this 5 min period. Analysis of variance of the pursuit rotor scores for the two groups during the last period failed to show any evidence of significance. In addition, reminiscence scores from the last trial of experimental period II to the first trial of experimental period III were computed and no significant difference was found between the two groups of subjects. It is obvious from the data that the two types of intervening activity, rest and P_d did not produce significant differences in subsequent pursuit rotor performance.

A word must be said about the significant differences found between trials during the initial practice period and experimental periods I, II and III, as well as the significant groups by trials interaction during experimental period I. Since scores on successive trials during these periods were not actually independent of each other, it is somewhat fallacious to ascribe the large numbers of degrees of freedom to the sources of variance. Therefore, a 'conservative test' has been devised (Greenhouse & Geisser, 1959) which reduces the degrees of freedom and thereby provides a more rigorous test of significance. When the 'conservative test' is employed, it is found that the differences between trials during the initial practice period, the between-trials differences during experimental period I, and the groups by trials differences during experimental period I are no longer significant. However, the differences between trials during experimental period II remain significant ($P < 0{\cdot}01$), while the differences between trials during experimental period III is significant ($P < 0{\cdot}05$), using the 'conservative test'. Our main conclusions, therefore, are not affected by the use of the conservative test.

DISCUSSION

The results of this experiment are unusually clear-cut. They are: (1) distraction depresses performance on the pursuit rotor; (2) the interference caused by distraction is a linear function of the amount of distraction provided; (3) distraction does not affect learning, but only performance; (4) distraction does not affect consolidation.

It may be useful briefly to discuss these results in the framework of consolidation theory. We may conceive of what happens during massed practice on the pursuit rotor, followed by a rest period, along the lines of the following diagram:

$$L \quad C_1 \quad M_1 \quad C_2 \quad M_2$$

In this diagram L stands for the original learning process which is set into motion during massed practice, and which may be conceived in some such way as the setting up of certain types of 'cell assemblies', for instance. These cell assemblies, however, are not immediately available for improved performance; they require a period of consolidation (C_1, or primary consolidation). Primary consolidation can only take place during a rest period and if no rest period is programmed within a given time after the original learning, the cell assemblies cease to function and are no longer available for consolidation. The period during which they are available is probably of the nature of 12–15 min or thereabouts. The length of time taken by primary consolidation to be complete depends of course on the amount of learning that has to be consolidated but apparently it does not exceed, in normal persons, 8 or 10 min (in schizophrenics there is evidence that it might take 24 hr or even more). If primary consolidation takes place it places the learning into short-term memory storage and thus makes it available for the improvement of performance. If the rest pause is continued there is now a transfer from short-term memory storage to long-term memory storage; this transfer is designated C_2 in the diagram and may be referred to as secondary consolidation. Its function essentially is to protect the memory from being disrupted by electric shock and other cerebral disturbances.

The distinction between primary and secondary consolidation is perhaps unusual; it is necessary because in the usual type of consolidation experiment reference is only made to secondary consolidation. This is due to the fact that practically all the work that has been done in this field has been done on tasks involving essentially spaced practice; primary consolidation occurs naturally under those conditions during the rest periods following each increment of learning. There are important differences, as well as important similarities, between primary and secondary consolidation; primary consolidation in the normal person is accomplished in a relatively short period of time, whereas secondary consolidation apparently takes considerably longer, possibly extending from 30 sec to several days. Unfortunately, no direct evidence is available on interference with secondary consolidation in skilled tasks performed under massed practice conditions, so that part at least of the theory outlined above must remain speculative.

It would appear that the cortical processes underlying primary consolidation, as well as those underlying secondary consolidation, are not easily disrupted by distraction, however severe, unless this distraction is very closely similar indeed to the activity itself which has been learned in the first place. The converse presumably also follows, i.e. the process of primary consolidation interferes with the practice of the original learning activity, if this is resumed before primary consolidation is completed, but does not for other types of activity (for one explanation of this interference see Eysenck, 1965).

We are left with the problem of learning. The theory outlined above accounts for the fact that there is very little improvement during massed practice; what little there is may be due to a certain amount of primary consolidation taking place during blocks, involuntary rest pauses, etc. Our results seem to clarify to some extent the nature of the conditions requisite for learning to take place. It does not appear that reinforcement, successful practice, or attention are closely concerned with the acquisition of skill during massed practice, as our groups showed significant differences in all these conditions during the practice period but failed to show any effects on learning. The only remaining condition which appears to be significantly related to learning is the actual exercise of the perceptual-motor actions required, and the drive conditions under which the individual is working (Eysenck, 1965). Clearly one experiment is not sufficient to settle so important a controversy (cf. also Briggs *et al.* 1957) but it is suggested that the use of distracting tasks may enable us to investigate the various theories mentioned at the beginning of this article in greater detail.

We are indebted to D.S.I.R. for the support of this investigation. Special thanks are due to Mr C. Attwood, Principal of Apprentice Training in the Ford Motor Co. at Dagenham, for his unstinting co-operation in arranging for the testing of apprentices at Ford's.

REFERENCES

BRIGGS, G, E., FITTS, P. M. & BAHRICK, H. P. (1957). Learning and performance in a complex tracking task as a function of visual noise. *J. exp. Psychol.* **53**, 379–87.

EYSENCK, H. J. (ed.) (1964). *Experiments in Motivation.* London: Pergamon.

EYSENCK, H. J. (1965). A three-factor theory of reminiscence. *Brit. J. Psychol.* **56**, 163–82.

GEISSLER, L. R. (1909). The measurement of attention. *Amer. J. Psychol.* **20**, 473–529.

GREENHOUSE, S. V. & GEISSER, S. (1959). On method in the analysis of profile data. *Psychometrika* **24**, 95–112.

KREEZER, G. L., HILL, J. O. & MANNING, W. (1954). Attention. *W.A.D.C. tech. Rep. no.* 54–155.

PAULI, R. (1930). Die Enge des Bewusstseins und ihre experimentalle Untersuchung. *Arch. ges. Psychol.* **74**, 201–57.

STERKY, K. & EYSENCK, H. J. (1965). Pursuit rotor performance as a function of different degrees of distraction. *Life Sci.* **4**, 889–97.

TUKEY, J. V. (1949). Comparing individual means in the analysis of variance. *Biometrics* **5**, 99–114.

(*Manuscript received* 19 *July* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 107–112
Printed in Great Britain

THE EFFECT OF DRIVE ON PERFORMANCE AND REMINISCENCE IN A COMPLEX TRACING TASK

By H. J. EYSENCK and R. A. WILLETT
Institute of Psychiatry, University of London

An experiment is reported in which high-drive and low-drive groups equated for intelligence were given a complex tracing task. Under conditions of spaced practice the low-drive group performed significantly better than the high-drive group, and similar differences were observed under conditions of massed practice. A rest pause of 10 min was interpolated in the performance of the groups tested under conditions of massed practice, and reminiscence was found to be greater for the low-drive groups than for the high-drive groups. Significant post-rest decline of performance under massed conditions was observed only for the groups having long pre-rest massed practice and not for those having short pre-rest massed practice.

One of the features of reminiscence and some of the phenomena attending it is the degree of task specificity observed (Eysenck, 1965). In a series of studies (Eysenck & Maxwell, 1961; Eysenck & Willett, 1961; Willett & Eysenck, 1962; Feldman, 1964) the writer and his colleagues have tested Kimble's (1950) hypothesis derived from Hullian theory, that in pursuit rotor learning high drive would lead to *better performance* and *greater reminiscence* than low drive; the results have verified the prediction in so far as it concerns reminiscence, but have failed to support the prediction regarding differences in performance. The present investigation was undertaken in an attempt to study the effects of drive on performance and reminiscence when a complex tracing task was substituted for the pursuit rotor. The task in question was chosen because it seemed possible to make predictions regarding the effects of drive on performance on the basis of Easterbrook's (1959) restatement of the Yerkes–Dodson Law in terms of cue utilization. He has proposed that 'the number of cues utilized in any situation tends to become smaller with increase in emotion', and that, in some tasks, 'proficiency demands the use of a wide range of cues, and drive is disorganizing or emotional'. The task chosen requires such use of a wide range of cues, and the prediction would therefore be made that high drive should lead to a decline in score. A preliminary study had supported this prediction (Eysenck & Willett, 1962).

A prediction about reminiscence is much more difficult to make because of the uncertain state of theory in regard to this concept. Eysenck (1965) has argued for a three-factor theory of reminiscence according to which the difference between pre-rest and post-rest performance on a task involving learning, such as pursuit rotor tracking or complex tracing as in this case, depends primarily on consolidation and not, as Hull and Kimble have argued, on the dissipation of reactive inhibition. If we assume that the amount of consolidation that takes place is a direct function of the amount of learning that has taken place during the pre-rest period, and if we assume that learning is impaired by high drive, then it would seem to follow that reminiscence should be greater in the low-drive than in the high-drive group. The assumption that high drive impairs learning as well as performance requires experimental support, of course, and may therefore be listed as a third prediction to be tested in the experiment itself.

METHOD

Subjects

The subjects were industrial apprentices aged 16–18, all male. The low-drive group were given the test under conditions such that the results were known by them not to affect their status in the company or their employment prospects. The high-drive group were given the test in a selection situation, and in such a manner that they were convinced that the results would affect their prospects of being taken on by the company, although in actual fact this was not so. Details about this method of producing differences in drive, and evidence regarding its effectiveness have been given elsewhere (Eysenck, 1964). All candidates were given a verbal intelligence test (N.I.I.P. Group Test 33), and high- and low-drive groups were equated within narrow limits on the scores of this test, so that no significant differences were found between main groups or subgroups. Such matching was considered necessary because correlations of between 0·4 and 0·5 were found between intelligence and performance in the various groups included in the experiment.*

Test

The test used was the Ammons (1955, 1960) version of the Pathways Test. There are twenty sheets to this test, each one of them having printed on it the numbers from 1 to 30 in a random arrangement, with a number 1 in the centre of the page. The subject has to trace these with a pencil, starting from the 1, and going on to the numbers in sequentional order; the score is the highest number reached in one minute. The twenty sheets were stapled together in random order so that each subject had a different sequence of sheets randomly arranged.

Experimental design

There were three conditions: *A*, *B* and *C*. In condition *A* there were 15 min massed practice, followed by a 10 min rest pause and 5 min of massed post-rest practice. In condition *B* there were 5 min massed practice followed by a 10 min rest pause, followed by 15 min of massed post-rest practice. Condition *C* was one of spaced practice, each 1 min trial being followed by 1 min of rest (Ammons, 1960, has shown that 1 min rest pauses were not inferior to 2½ or 5 min rest pauses). During the rest periods in all conditions subjects were prevented from looking at test pages. The number of subjects in the high drive groups (*A*, *B*, *C* in order) was 71, 75 and 67; and in the low drive groups 86, 80 and 81.

RESULTS

Fig. 1 shows the results for condition *C*: the high-drive group scored well below the low-drive group at all points. There was also a greater improvement in performance for the low-drive group than for the high-drive group, suggesting that high drive on this task does interfere with learning. Analysis of variance (Table 1) confirmed the significance of the difference between drives ($P < 0{\cdot}001$); differences between trials were also significant ($P < 0{\cdot}001$), while the interaction between drives and trials is significant ($P < 0{\cdot}05$). As trials are not independent of each other the conservative test was applied in reaching these values (Greenhouse & Geisser, 1959), and in all other analyses in this paper.

---

* In previous work (e.g. Eysenck, 1964), groups were not matched for intelligence and this point has at times been criticized. There are several reasons why such a matching was not considered necessary. (1) Many of the tasks used, such as pursuit rotor learning, are not correlated with intelligence so that matching would be unnecessary. (2) For the great majority of tasks higher performance was predicted, and found, for the high-drive group; elimination of the less intelligent from this group to equate it with the low-drive group would have enhanced the predicted effect, but there seemed to be no need for this to be done as the results were already significant enough. (3) Even where prediction favours the low-drive group, and where the test in question correlates quite highly with intelligence, as in the test used in this paper, correction for intelligence makes surprisingly little difference to the outcome as will be seen by comparing the results here reported, where intelligence was controlled, with those of the Eysenck & Willett (1962) study where it was not controlled. On the whole, criticisms of the studies reported in Eysenck (1964) on the grounds that intelligence was not controlled would not seem to be well taken, with the possible exception of the experiments relating to digit span, pages 137–142.

Table 1. *Analysis of variance for condition C, spaced practice*

Source	D.F.	Sum of squares	Mean squares	*F*	*P*
Between drives	1	5857·26	5857·26	54·13	< 0·001
Between trials	19	2824·75	148·67	10·83	< 0·001
Drives × trials	19	1158·15	60 95	4·44	< 0·05
Subjects within drives	145	15689·13	108·20	—	—
Residual	2755	37821·60	13·73	—	—
Total	2939	63350·89	—	—	—

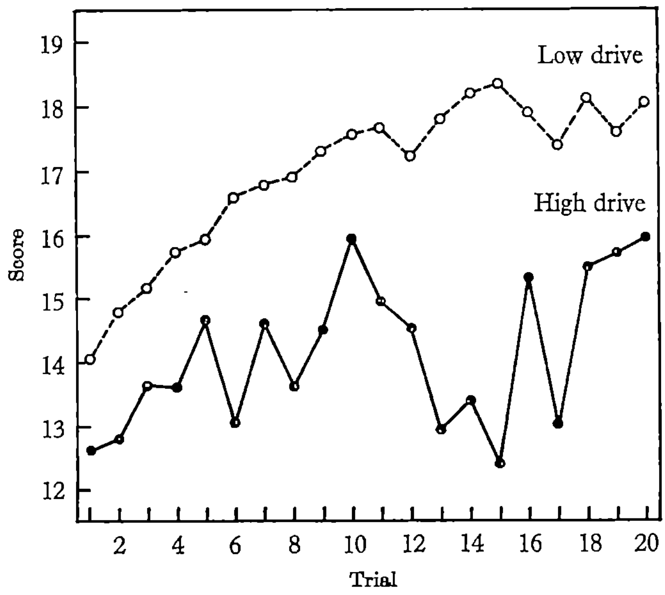


Fig. 1. Scores of high drive and low-drive groups for twenty spaced practice trials.

Fig. 2 shows the results of condition *A*: again the low-drive group was superior at all points to the high-drive group and, as predicted, it showed a greater reminiscence score. Analysis of variance of pre-rest scores showed the difference between drives to be significant ($P < 0{\cdot}01$), while no significance attached to trial differences; there was thus no evidence of learning under conditions of massed practice. During post-rest trials the difference between drives is very much more significant than before ($P < 0{\cdot}001$) and the difference between trials achieved significance ($P < 0{\cdot}05$), performance declining as is usual during post-rest practice under massed conditions.

Fig. 3 shows results for condition B: here too the low-drive group was superior at all points to the high-drive group, and reminiscence was greater for the low-drive than for the high-drive group. Analysis of variance showed level of performance during pre-rest to be significantly different ($P < 0{\cdot}05$), with no significance attaching to the trials or interaction terms; there was again no evidence of improvement under conditions of massed practice. Post-rest performance shows a difference between drives significant ($P < 0{\cdot}001$), with no significance attaching to the trials or interaction terms.

The results of analysis of variance on the four reminiscence scores are given in Table 2 (*b*); Table 2 (*a*) gives the mean reminiscence values of the four groups. Long

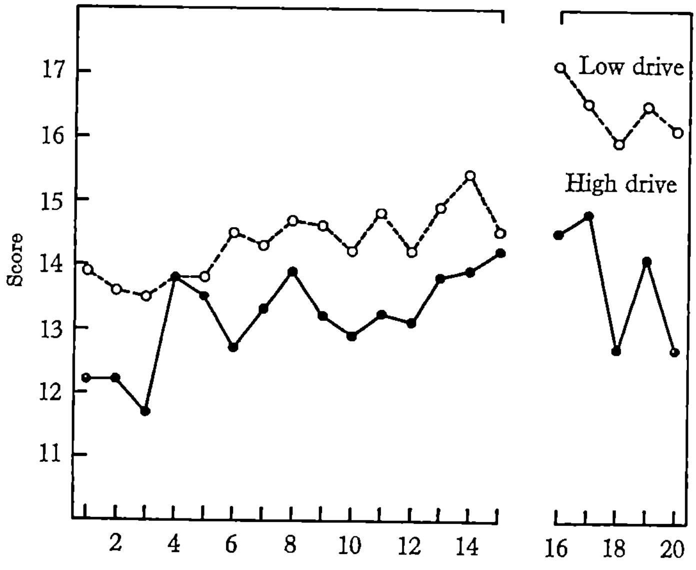


Fig. 2

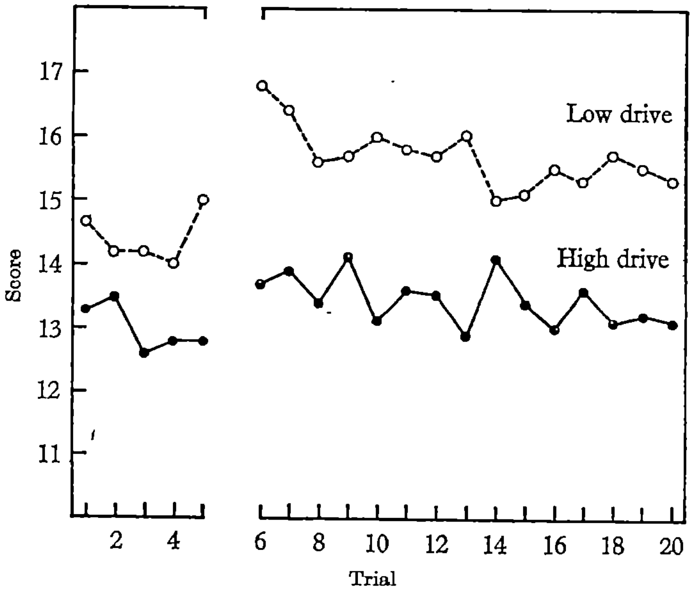


Fig. 3

Figs. 2 and 3. Scores of high- and low-drive groups during trials preceding and following a 10 min rest pause.

pre-rest resulted in slightly higher reminiscence, but this difference was non-significant. Low drive resulted in significantly greater reminiscence than high drive. The interaction was non-significant.

Table 2

(*a*) Mean reminiscence scores

	Drive		
Pre-rest	High	Low	
Long	0·282	2·570	1·535
Short	0·960	1·738	1·361
	0·630	2·169	

(*b*) Analysis of variance of reminiscence scores.

Source	D.F.	Sum of squares	Mean squares	F	P
Between drives	1	183·78	183·87	6·1516	< 0·05
Between length pre-rest	1	2·35	2·35	—	
Drives × pre-rest	1	43·14	43·14	1·4433	N.S.
Residual	308	9205·82	29·89		
Total	311	9435·18			

DISCUSSION

The results of the experiment seem to be in good accord with prediction. High drive impairs performance under all conditions, i.e. whether spaced or massed (Eysenck & Willett, 1962); high drive impairs learning as evidenced by the spaced condition of practice; and high drive, possibly in consequence of this interference with learning, produces significantly less reminiscence. In all these features results differ from those reported on the pursuit rotor, thus demonstrating again the importance of task specificity.

One or two other features of the results may be worthy of comment. Ammons (1947) has postulated the existence of permanent work decrement as a function of length of pre-rest practice, and when we compare groups *A* and *B* on trials 16 and 6 respectively with group *C*, it will be seen that there is evidence for such a permanent decrement in group *A* but not in group *B*. This effect would be explained by Kimble (1950) and Ammons (1960) in terms of conditioned inhibition, but of course our data throw no light on the actual mechanism responsible for the effect. In any case, the decrement is not statistically significant.

The post-rest decline in performance under conditions of massed practice has been explained by Eysenck (1965) as possibly being due to the on-going and interfering effects of consolidation. If this hypothesis were true then we would expect greater interference effect in group *A* than in group *B* because group *A* had been practising for a longer time and would therefore have more material to consolidate, thus producing greater interference. As we have noted before, this difference is in fact found, decrement being significant under post-rest only for group *A* and not for group *B*.

We have argued that in terms of our theory the lower reminiscence scores for the high-drive group are the effect of less learning going on during the pre-rest massed practice period. It might be argued that what has happened is an interference with practice during this period which results in lower performance and, therefore, inevitably in less learning. Eysenck & Thompson (1966) have shown, however, that interference with practice does not necessarily produce differences in learning; in their work an interfering task produced decrement of performance on the pursuit rotor but the effects of this distracting task were completely eliminated by the rest period, so that, in fact, the groups showed no difference in learning. This suggests, even when allowance is made for the difference in task, that drive affects learning directly rather than through interference with performance.

The work was supported by D.S.I.R. We are indebted to Mr C. Attwood, Principal of Apprentice Training in the Ford Motor Co. at Dagenham, for his unstinting co-operation in arranging for the testing of apprentices.

REFERENCES

AMMONS, C. H. (1955). Tasks for the study of perceptual learning and performance variables. *Percept. mot. Skills* **5**, 11–14.

AMMONS, C. H. (1960). Temporary and permanent inhibitory effects associated with acquisition of a simple perceptual-motor skill. *J. gen. Psychol.* **62**, 223–45.

AMMONS, R. B. (1947). Acquisition of motor skill. I. Quantitative analysis and theoretical formulation. *Psychol. Rev.* **54**, 263–81.

EASTERBROOK, J. A. (1959). The effect of emotion on cue utilisation and the organisation of behaviour. *Psychol. Rev.* **66**, 183–201.

EYSENCK, H. J. (ed.) (1964). *Experiments in Motivation.* Oxford: Pergamon.

EYSENCK, H. J. (1965). A three-factor theory of reminiscence, *Brit. J. Psychol.* **56**, 163–82.

EYSENCK, H. J. & MAXWELL, A. E. (1961). Reminiscence as a function of drive. *Brit. J. Psychol.* **52**, 43–52.

EYSENCK, H. J. & THOMPSON, W. (1966). The effects of distraction on pursuit rotor learning, performance and reminiscence. *Brit. J. Psychol.* **57**, 99–106.

EYSENCK, H. J. & WILLETT, R. A. (1961). The measurement of motivation through the use of objective indices. *J. ment. Sci.* **107**, 961–8.

EYSENCK, H. J. & WILLETT, R. A. (1962). Cue utilization as a function of drive: an experimental study. *Percept. mot. Skills*, **15**, 229–30.

FELDMAN, M. P. (1964). Pursuit rotor performance and reminiscence as a function of drive level. In *Experiments in Motivation*, ed. H. J. Eysenck. Oxford: Pergamon.

GREENHOUSE, S. W. & GEISSER, S. (1959). On method in the analysis of profile data. *Psychometrika* **24**, 95–112.

KIMBLE, G. A. (1950). Evidence for the role of motivation in determining the amount of reminiscence in pursuit rotor learning. *J. exp. Psychol.* **40**, 240–53.

WILLETT, R. A. & EYSENCK, H. J. (1962). An experimental study of human motivation. *Life Sci.* **1**, 119–27.

(*Manuscript received* 27 *July* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp*. 113–119
Printed in Great Britain

AGE AND PERCEPTUAL-MOTOR TRANSFER IN IMBECILES: TASK COMPLEXITY AS A VARIABLE

BY A. D. B. CLARKE AND G. M. COOPER
Department of Psychology, University of Hull

Earlier work with imbecile subjects suggested a relation between low chronological age and amount of transfer, in the manner suggested by McGeoch and Hebb. It remained possible, however, that adults were nearer their ultimate limits and had benefited from the transfer processes earlier in life. It was necessary therefore to evolve a method enabling a fair comparison to be made between children and adults. Task complexity was thus adjusted either for adults or for children so that identical starting scores were obtained. Two experiments yielded similar acquisition curves and transfer scores for both groups, suggesting that task complexity is a major variable in transfer. Hebb's view that adults will have already gained maximum benefit from transfer early in life may only hold for simple or oft-used processes. Transfer seems to be a function of early development indirectly rather than directly, for all experiences will tend to be complex and unfamiliar to the young organism.

In an earlier paper, Clarke & Blakemore (1961) investigated the hypothesis that transfer of training was inversely proportional to age, using imbecile subjects of different ages. The hypothesis was confirmed on various discrimination sorting tasks, 9-year-old subjects showing very great transfer in comparison with those aged 17 and 23. A second paper (Clarke & Cookson, 1962) reported further investigations of the learning processes of the same 9-year-old subjects. Similar discrimination sorting tasks were again employed, but with problems of greater complexity. Once again it was found that the young subjects made greater transfer gains than adults on the more complex material. Moreover, the retention of learning and its facilitatory effect on re-learning or new learning was demonstrated over 6-month and 1-year intervals. An attempt was made to identify some of the parameters of 'learning sets', and the hypothesis was raised that the main deficit in imbecile discrimination learning lies in acquisition or retention, but rather in poor starting-points. This suggested a relative deficiency in spontaneous learning from ordinary life experience.

In summary, all these experiments demonstrated the ease with which transfer takes place in tasks similar in type and difficulty but different in content, using young imbecile subjects. The idea that transfer was a main function of early development, first envisaged by McGeoch and by Hebb, thus received support. Indeed, it seemed relatively difficult, using industrially trained adult imbeciles, to demonstrate transfer as an impressive function, yet relatively easy with children.

The researches mentioned above raised two main problems. First, it seemed possible that the consistently smaller transfer shown by adolescent and adult subjects may have been due to their initially superior performance and thus their closer initial approximation to ultimate limits. It could also well be that transfer had to some extent occurred earlier in life, thus being measured after most of it had already taken place. Secondly, although the task pairs possessed different contents, these did not represent totally different material so that no data existed indicating the width or narrowness of the process. The present paper offers a solution to the first of these two problems; data on the latter are given in Clarke, Cooper & Henney (1966).

A method was needed which would enable more exact comparisons to be made between children and adults. There seemed to be two ways in which these could be placed on an equal footing: either (*a*) to accept the children's starting time as the criterion, and increase task complexity for the adults until approximately the same value was reached, or (*b*) to take the initial adult performance as the criterion and decrease task complexity for the children until the two were matched. In both cases the experimental material had to be either identical or closely similar in its demands and characteristics. With the two very different age-groups now at the same point in learning, it should be possible to establish whether or not transfer was a direct function of low chronological age.

As in our previous work, visual discrimination tasks were used, pairs of tasks being matched for difficulty on the basis of total time to complete an error-free performance, using self-correction. The following transfer paradigm was employed:

Group I subjects	Pre-test task B (1 trial)	Learn task A (10 trials)	Learn task B (10 trials)
Group II subjects	Pre-test task A (1 trial)	Learn task B (10 trials)	Learn task A (10 trials)
	Sum of scores	Learning I	Learning II

Pre-tests took place on a Friday, the first 10 learning trials being given over the subsequent week (2 trials per day) and the second 10 trials over the succeeding week. Each group was matched with the other on the basis of initial score, and thereafter acted as the other group's control.

Experiment I: typewriter key sorting, using starting score of children as the criterion for adults

Method

Subjects. In the first experiment only one group (adults) was used *de novo*, the results being compared with those previously obtained for children (Clarke & Blakemore, 1961). Twenty-five industrially trained adult imbeciles, working in the Hull Training Centre were used, their mean I.Q. being 44 and mean age 23 years.

Tasks. Two new parallel tasks were produced for the adults, scaled for difficulty on the basis of the original children's scores. These involved the discrimination of seven rather than five symbols. Fig. 1 shows the material originally used for the children and the new task designed for adults. The procedure adopted was identical with that used previously: for each task seven small tins were arranged in a semicircle on a table in front of the subject; on each was mounted a different typewriter key and in front of them was a pile of thirty-five keys placed in random order. The keys had to be sorted correctly, with five keys for each tin.

Instructions. A minimum of verbal instructions was used. Each subject was put at ease in the experimental situation and then shown the material. A demonstration of the placement of two of the stimuli was given, after which these particular keys were replaced in the pile. The subject was then asked 'Can you do that? Carry on.'

Results

Fig. 2 shows that a closely matched starting time had been achieved on pre-test, and that the general shape of the learning curves is very similar. Table 1 shows the results treated statistically. Transfer was measured as the time difference between each of ten trials of learning I and the ten trials of learning II.

Cross-over analysis of variance (Maxwell, 1958) for the new group of adult subjects indicates (*a*) a highly significant transfer effect from learning I to II ($P < 0\cdot001$); and (*b*) a significant difference between the two subgroups of adults ($P < 0\cdot01$), which was controlled for in the experimental design by alternating the task order

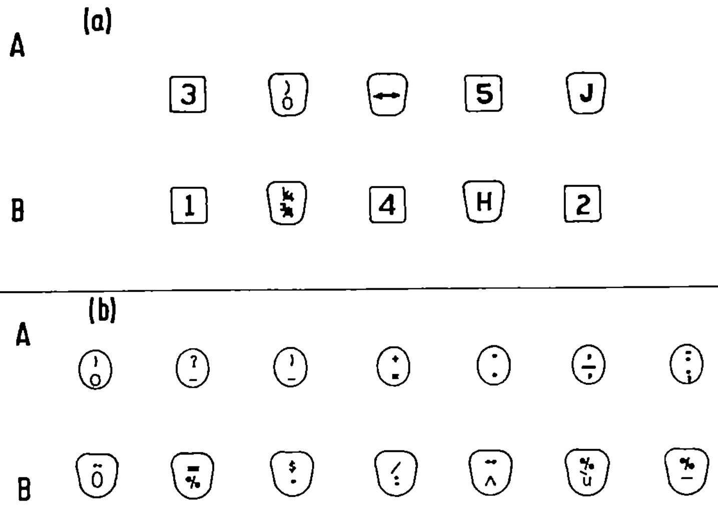


Fig. 1. Pairs of typewriter key-sorting tasks (*a*) for children and (*b*) adults, designed to achieve equal relative complexity for both groups. Each key is mounted on a tin, and five other keys bearing the same symbol have to be sorted from a pile on the table into each tin. The extra number and difficulty of the discriminations needed for the adult task increased their sorting times to those of the children on the simpler task.

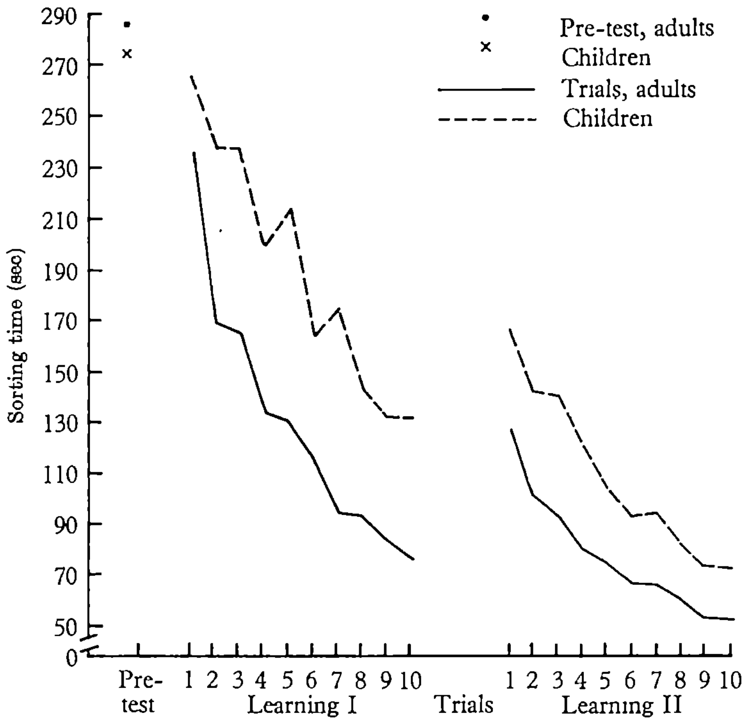


Fig. 2. Mean time-scores of adults and children on the learning of pairs of typewriter key-sorting tasks. Relative task complexity was held constant for adults and children who were roughly matched on pre-test. The similarity in the acquisition curves, and hence in transfer, is striking.

so that for half the subjects task A preceded task B, and conversely for the other half (see transfer paradigm). It was thought inadvisable to use cross-over analysis of variance on the earlier data for children because of the small numbers. The former experimental design, however, was identical to the present one.

Table 1. *Key sorting tasks: analysis of variance for adult imbeciles*

Source of variation	D.F.	Mean square	F	P
Between learning I and learning II	1	3353333	43·938	< 0·001
Between tasks	1	215338	2·822	N.S.
Between groups	1	697213	9·135	< 0·01
Residual	22	76319	—	—

The scores of one subject, chosen at random, were omitted from the analysis to achieve equal numbers in the groups.

The earlier data on the children showed highly significant transfer from learning I to II ($t = 13{\cdot}47$, D.F. 9, $P < 0{\cdot}001$). The difference between the transfer scores of children and adults now yields $t = 5{\cdot}22$, D.F. 9 ($P < 0{\cdot}01$), indicating that the children showed significantly greater transfer than the adults. Even so, as shown by Fig. 2, the similarity of the acquisition curves is very striking and unexpected. It indicates at least that, when complexity was roughly equated, adults and children were far less dissimilar in their performance of these tasks than the earlier work had suggested. The results of this first experiment may also have been affected by differences in populations (hospital versus training centre) and the relatively greater sophistication of industrially trained adults who were living in the community.

Experiment II: using adult imbecile starting score as criterion for children

Method

Minnesota task. A number of experiments have used the Minnesota Spatial Relations Test as a learning task (e.g. Tizard & Loos, 1954). This consists of four large boards, A, B, C and D, each of which contains fifty-eight holes into which corresponding forms have to be fitted. The forms differ in shape, size and orientation, and one set is common to boards A and B, and another to C and D, although within each pair of boards the location of the holes is quite different. For present purposes, forms from board A had to be placed in B, and forms from board C had to be placed in D. Alternate subjects commenced with alternate formboards. Groups were again matched on the basis of initial pre-test score. These are difficult tasks both for children and for adult imbeciles. In order to reduce the starting time of the children to that of adult subjects, about one-third of boards B and D were masked and the corresponding forms from A and C removed (nineteen and eighteen pieces, respectively). In this way it was hoped to equate the starting points of children and adults and thus to give them virtually identical tasks of equal complexity (defined in relative terms).

Procedure. Instructions and procedures were similar to those of Expt. I. All subjects had to place the pieces in the boards as quickly as possible, after an initial single demonstration of the placement of two forms. These were always arranged in a constant array on the far side of the board in use. Conditions of encouragement and reward were the same as earlier; alternate subjects commenced on alternate tasks and after a single pre-test were given one trial per day for 8 days, followed by a further eight trials on the alternate board.

Results

Fig. 3 shows that very closely similar curves result from equal starting points on the same tasks for both adults and children. Table 2 shows the results in statistical

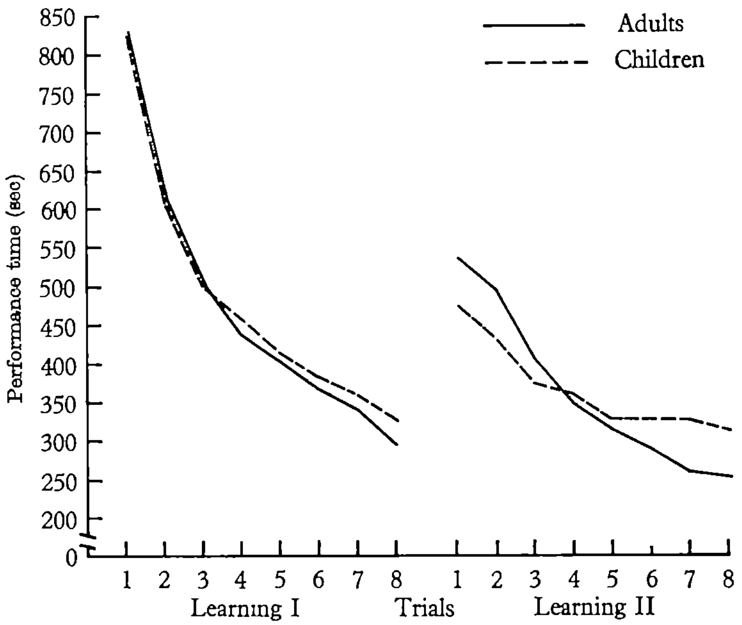


Fig. 3. Mean time-scores of adults and children on Minnesota formboard tasks. By masking about one-third of each board the times taken by the children were reduced almost exactly to those for trial 1 by adults on the full formboard. Acquisition and transfer functions are almost identical.

form. Once again, differences in the time taken on each of the eight trials of learning I and the eight trials of learning II were used as a measure of transfer.

As expected, both adults and children show highly significant transfer from learning I to II. The difference between tasks C–D and A–B is also highly significant, the former being more difficult, but this is of course controlled in the experimental design. The difference between groups is significant for the adults and to a less extent for the children, as is to be expected in view of the large inter-task differences and consequent differential transfer effects.

In Expt. II, as might be surmised from Fig. 3, the difference between transfer scores for adults and children is not significant ($t = 0{\cdot}4$). We can conclude therefore that, using similar (training centre) subjects and tasks common to adults and

Table 2. *Minnesota formboard tasks: analyses of variance of 14 adult and 12 child imbeciles*

Source of variation	D.F.	Mean square	*F*	*P*
	(*a*) Adults			
Between learning I and learning II	1	7127576	34·585	< 0·001
Between tasks	1	4436107	21·525	< 0·001
Between groups	1	2320704	11·261	< 0·01
Residual	12	206090	—	—
	(*b*) Children			
Between learning I and learning II	1	5317592	66·425	< 0·001
Between tasks	1	6265882	78·271	< 0·001
Between groups	1	595035	7·433	< 0·05
Residual	10	80054	—	—

children, there is no difference in the considerable amount of transfer shown by them when their starting-points are identical, that is, when relative complexity is held constant.

DISCUSSION

The results of the experiments suggest that task complexity is an important variable in learning transfer for this population. Indeed, a re-examination of transfer in relation to the varying difficulties of the different tasks shows the same principle operating in the experiments reported by Clarke & Blakemore (1961) and Clarke & Cookson (1962); that is, the greater the difficulty (as measured by time taken on the first trial) the greater the transfer. Nevertheless, the present finding was partly anticipated in the view that 'in the adolescent or adult, transfer of training will only be considerable in tasks which are very new to the subject, or which demand the exercise of relatively unexercised functions' (Clarke & Blakemore, 1961).

The hypothesis of McGeoch and of Hebb that adults are likely to have gained maximum benefit from learning transfer earlier in life is thus not supported, although it may well hold for simple or oft-used processes. Transfer may, however, be associated *indirectly* with low chronological age in the sense that, to the young organism, all tasks will tend to be complex, so that transfer early in life may well be a very common phenomenon.

In asking what has been transferred from task to task, Clarke & Cookson (1962) offered evidence for three interlinked variables underlying the vague notion 'learning set'. These may be termed task expectancy, improved perceptual discrimination (perhaps akin to 'acquired distinctiveness of cues') and improved conceptual discrimination. In all these, both retention of earlier relevant experience and a general 'sharpening' of perceptual and conceptual processes are implied. It did not then appear, and still does not, that increased motor speed *per se* is in any way relevant. The present experiments again seem to underline the earlier evidence. Subjects in both types of task tended gradually and spontaneously to abandon the sorting of keys or shapes at random, moving to strategies of dealing with a single category at a time selected from among all the others.

Task complexity, considered always in relative terms, arises from several variables. In the first experiment, for example, the number of items to be discriminated and the number of categories are relevant, as are the perceptual qualities of the stimuli, i.e. their size, random placement and the degree of similarity between different symbols. In the second experiment, the obvious complexity variables are the area of the form boards, the number of shapes and the number of shape categories, their shape, size orientation and number, and the degree of similarity between different categories. In discrimination learning, task difficulty has usually been considered in terms of easy–hard, or hard–easy sequences, and the evidence concerning these complicated relationships has been discussed by Lawrence (1952), House & Zeaman (1960) and by Holding (1962). In the case of all our experiments, however, the tasks involve hard–hard sequences for our subjects, and apart from task expectancy and a motor movement of extreme simplicity it is to be noted that no identical stimulus elements were involved. It thus appears that tasks must be complex for transfer to be demonstrated, and that if direct comparisons of learning and transfer are to be made between subjects of different ages, or different abilities, then some experimental adjustment of relative task complexity needs to be incorporated in the design. Current work yields additional support for this thesis, together with evidence on width of transfer (cf. Clarke *et al.* 1966).

The writers are very grateful to the Nuffield Foundation for generous financial support; to Dr A. Hutchison, O.B.E., Medical Officer of Health, Kingston-upon-Hull, and Dr W. Fairlamb, Senior Medical Officer for Mental Health, and their staff, and to Dr G. W. Waters and his staff at Brandesburton Hospital, for excellent research facilities; to Mrs E. Cooper for assistance with one of the experiments; to Dr A. M. Clarke, to Mr J. P. N. Phillips and Mr A. B. Royse for helpful advice.

REFERENCES

CLARKE, A. D. B. & BLAKEMORE, C. B. (1961). Age and perceptual-motor transfer in imbeciles. *Brit. J. Psychol.* **52**, 125–31.

CLARKE, A. D. B. & COOKSON, M. (1962). Perceptual-motor transfer in imbeciles: a second series of experiments. *Brit. J. Psychol.* **53**, 321–30.

CLARKE, A. M., COOPER, G. M. & HENNEY, A. S. (1966). Width of transfer and task complexity in the conceptual learning of imbeciles. *Brit. J. Psychol.* **57**, 121–28.

HOLDING, D. H. (1962). Transfer between difficult and easy tasks. *Brit. J. Psychol.* **53**, 397–407.

HOUSE, B. J. & ZEAMAN, D. (1960). *Learning and Transfer in Mental Defectives.* University of Connecticut, Progress Report no. 2.

LAWRENCE, D. H. (1952). The transfer of a discrimination along a continuum. *J. comp. physiol. Psychol.* **45**, 511–16.

MAXWELL, A. E. (1958). *Experimental Design in Psychology and the Medical Sciences.* London: Methuen.

TIZARD, J. & LOOS, F. M. (1954). The learning of a spatial relations test by adult imbeciles. *Amer. J. ment. Defic.* **59**, 85–90.

(*Manuscript received* 6 *September* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 121–128
Printed in Great Britain

WIDTH OF TRANSFER AND TASK COMPLEXITY IN THE CONCEPTUAL LEARNING OF IMBECILES

By ANN M. CLARKE, G. M. COOPER and A. S. HENNEY
Department of Psychology, University of Hull

This paper offers further evidence on the relation between task complexity and amount of transfer, using imbecile subjects. The main aim of the investigation was to determine whether transfer might occur between widely different tasks, requiring some degree of conceptual behaviour for their performance. The transfer task consisted of photographs of objects within five common categories; three different training tasks of varying complexity required the sorting of geometrical shapes. The two more complex of these training conditions resulted in considerable transfer on retest with the photographic material, and upon its subsequent direct acquisition. The simple geometric training gave results mid-way between the former and those of a control group. Transfer between widely different tasks has thus been demonstrated and the possible implications for understanding of the development of conceptual abilities are briefly considered.

The work that is reported arises directly from earlier studies by Clarke and his colleagues (Clarke & Blakemore, 1961; Clarke & Cookson, 1962; Clarke & Cooper, 1966). Starting with the hypothesis that transfer might be a main feature of low chronological and hence mental age, the conclusion was reached in the most recent paper that, except for simple and oft-used processes, age was probably only indirectly relevant and that a main factor influencing amount of transfer was the complexity of the intervening training task. In each of these experiments, transfer and training tasks were similar in nature and experimental demands, though different in content (e.g. two different formboards of similar difficulty comprised one such task pair).

Implicit in most work on transfer is the assumption that there must be an obvious similarity between the tasks used for positive transfer to occur. In experiments using visual material, the similarity is commonly to be found in the stimulus components; where training takes the form of instruction in principles, these are clearly related to the transfer task (Judd, 1908; Hendrickson & Schroeder, 1941); in the field of verbal behaviour once again similarity is normally to be found either in the stimulus materials used, or in the associations they evoke. Thorndike's attack on the theory of formal discipline and his propositions regarding identical elements as a necessary attribute of positive transfer have been rather rigidly interpreted by research workers. No attempts appear to have been made to explore systematically the possibility of transfer occurring through modification of a mediational response system. This is likely to be demonstrable primarily in young and inexperienced subjects who have not had the opportunity spontaneously to acquire the efficient mediational systems which are normally present at later stages of human development. Since imbeciles are by definition particularly poor abstractors and are deficient in spontaneous learning ability, it was considered that they might be particularly suitable subjects for the investigation of this problem.

In summary, the aims of the present investigation were to discover (*a*) whether the factor of complexity would again be shown to be relevant, using a different experimental design and holding age constant, (*b*) whether evidence would be gained on modification of mediational systems by using material in transfer and training tests which was widely different, and (*c*) whether any differences in the amount of transfer induced by training of different degrees of difficulty would subsequently affect the rate or level of learning of transfer material. As will be appreciated, such questions might have a bearing on the rôle of transfer in the development of cognitive processes.

Difficulty and complexity are not seen as necessarily synonymous attributes, and they can obviously vary along several dimensions. The decision as to how the tasks were to vary in complexity was to a large extent arbitrary, and does not necessarily represent the best method. Some evidence on this point emerged during the course of the experiment.

Two main hypotheses were formulated: (1) that transfer can be demonstrated where the training and transfer tasks are widely different both perceptually and conceptually, and where both involve a higher degree of abstraction than occurs in matching tasks; (2) that amount of transfer would be related to the level of difficulty of the intervening training task. It is assumed that no transfer will occur when the training task is either so easy or so difficult that no learning can occur.

Two measures of transfer were used: (*a*) performance on immediate retest contrasted with pretest, and (*b*) the level of subsequent direct acquisition of the transfer task following the retest.

METHOD

Experimental design. Four groups of subjects were equated on the basis of performance on pretest in the transfer task, T, and on a number of other variables. Thereafter the four groups were randomly allocated to four conditions, training I, II, III and control, for twenty-four spaced trials. After retest on T, nine direct learning trials on this material were given, making a total of eleven trials on task T for all four groups.

The tasks. All tasks involved sorting twenty response cards, 3 in. × 3 in., into five categories represented by stimulus cards.

The transfer task T consisted of black and white photographs of objects belonging to five common but rather wide conceptual categories: human beings, animals, furniture, tableware and clothes (see Plate 1). It was of special interest to discover how imbecile subjects, whose language development is particularly poor, would deal with this material without the specific aids of verbal mediation.

Training task I (easy) involved straight matching of twenty large shapes filled in with achromatic content. See top line of Plate 2, which represents the stimulus cards, for which response cards (in this case not shown) were identical in shape, size, orientation and content.

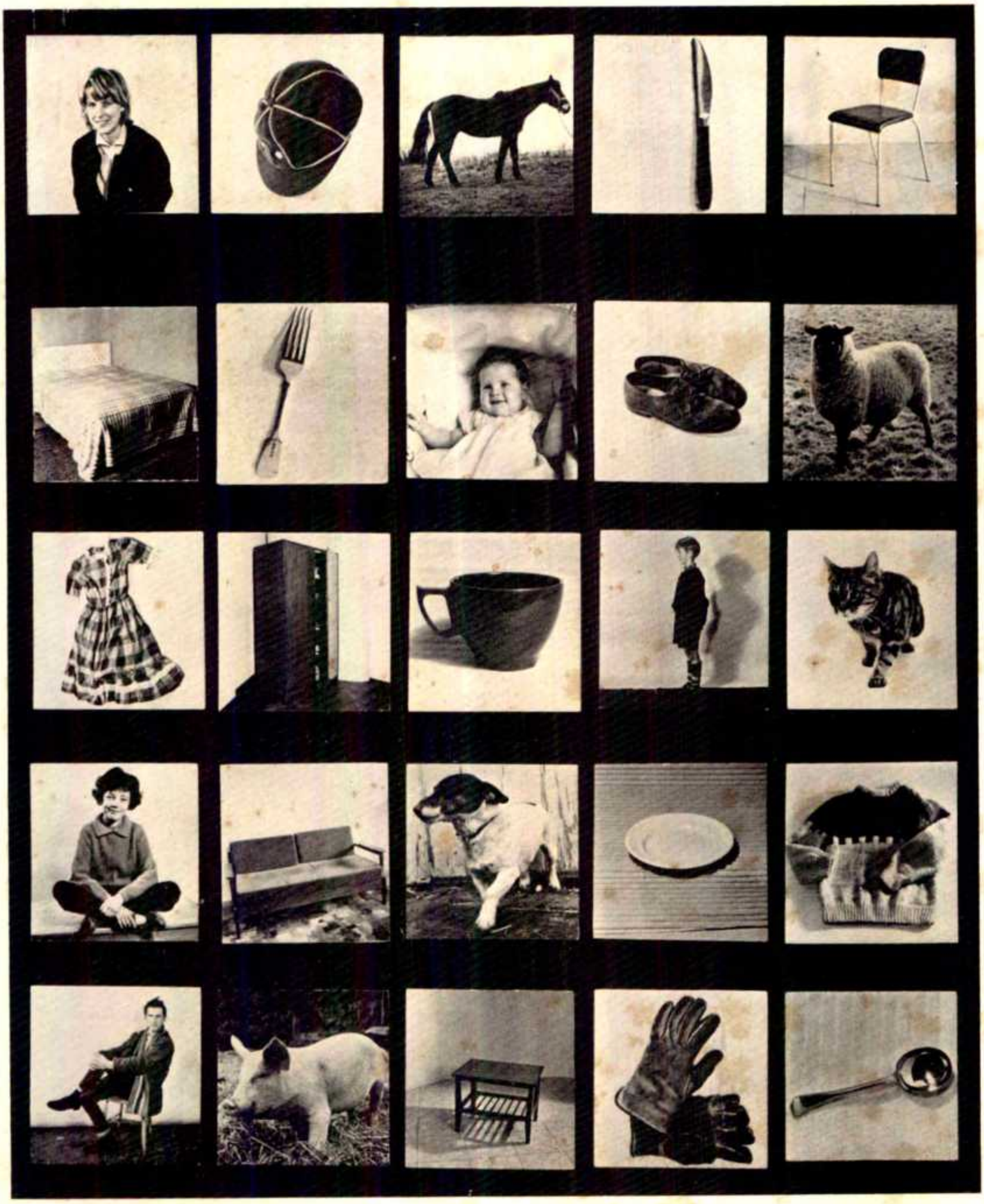
Training task II (fairly difficult) involved sorting the same shapes as in I but without achromatic content. In this case, however, the designs on the response cards differed in size and orientation from those on the stimulus cards. (The *outlines* only of the material shown in Plate 2 were used.)

Training task III (difficult) is shown in Plate 2. The top line represents the stimuli and the bottom four lines the response cards in the order given from left to right. Subjects were required to sort for shape, ignoring content which thus operated as a distractor.

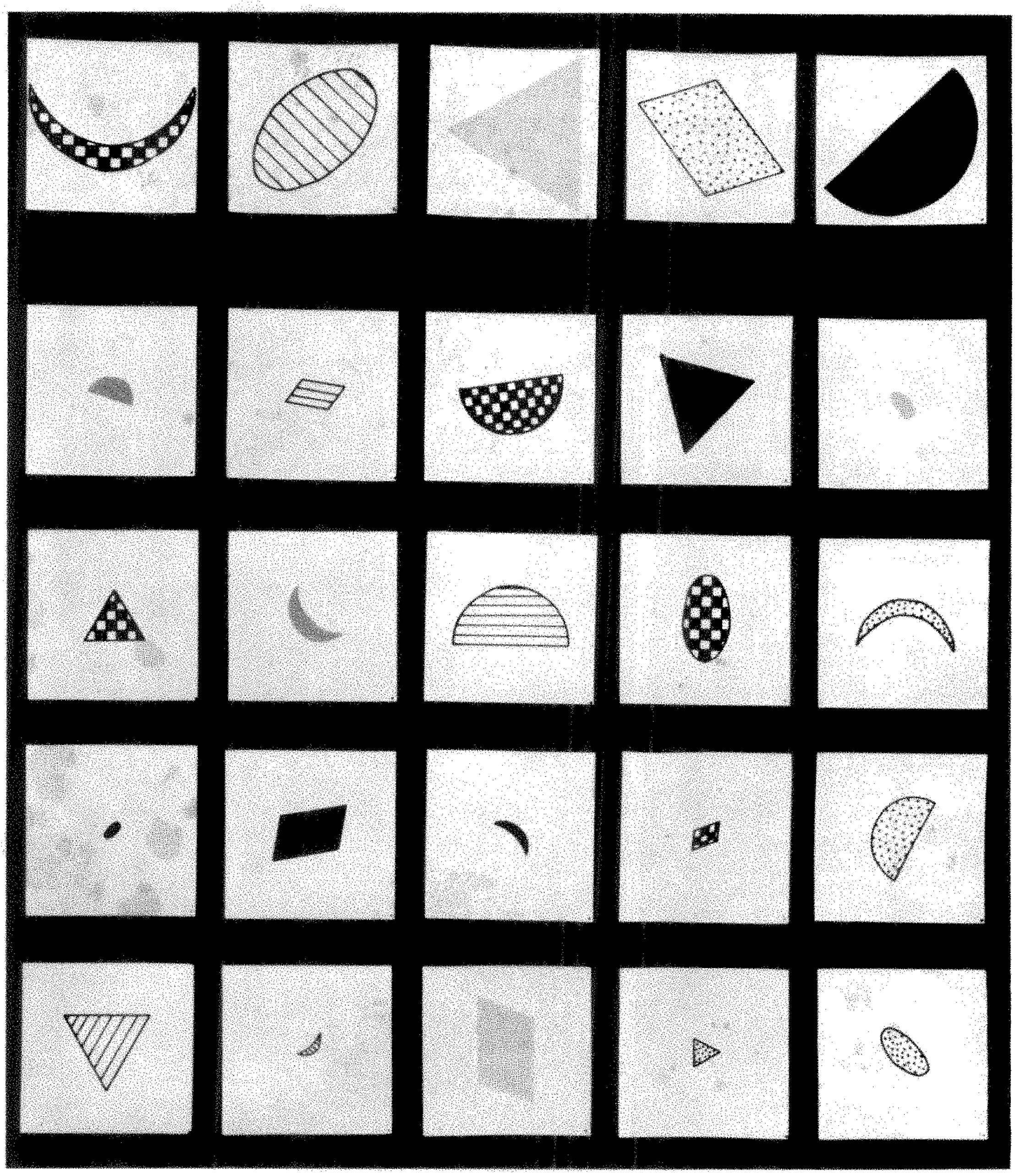
Control conditions involved attendance of subjects on alternate days in the experimental room for conversation and social contact with the experimenters.

Subjects. These were thirty-six imbeciles drawn from a local day training centre and a residential institution. They were divided into four groups of nine subjects and the following factors were controlled: age, I.Q., number of mongols per group, place of residence (home or institution), and pretest score on the transfer task T.

The groups were randomly assigned to type of treatment. Table 1 gives a summary of the initial data on them. No significant differences were found among the initial means, and the variances were homogeneous.



Transfer task. Top line indicates stimulus cards, and response cards below have to be sorted into the five categories. Order of presentation, top left to bottom right.



Training task III. Top line indicates stimulus cards, and response cards below have to be sorted into the five categories. Order of presentation, top left to bottom right. Subjects were required to sort for shape, ignoring content.

ANN M. CLARKE, G. M. COOPER AND A. S. HENNEY

Procedure. The apparatus consisted of a panel holding the five stimulus cards, each with a box below, the floor of which could be lowered by levers at the side. A red light bulb was fixed in each box immediately beneath the stimulus card.

Each subject was put at ease on entering the experimental room. He was seated in front of the apparatus and verbal instructions were kept to a minimum. Initially two cards, not subsequently used in the experiment, were slowly placed in their correct compartment, the red light flashed and the card dropped out of sight. The subject was then told, 'I want you to put all the pictures I give you into their proper places'.

Table 1. *Basic data on groups*

	Group							
	II		III		I		Control	
n...	9		9		9		9	
Mean age (yr)	12·5		12·1		12·5		12·4	
Mean I.Q.	32·9		34·6		37·0		36·1	
Mongols/group	4		4		4		4	
Transfer task T	Sorting score*	Errors	Sorting score*	Errors	Sorting score*	Errors	Sorting score*	Errors
Pretest mean	8·78	29·00	8·33	25·55	8·11	27·77	8·56	24·33
Retest mean	14·11	12·56	12·33	14·33	10·44	22·10	7·34	27·44
Difference	+5·33	−16·44	+4·00	−11·22	+2·33	−5·67	−1·22	+3·10
Training task								
First trial mean	15·0	11·0	13·0	15·0	19·3	0·8	—	—

* The maximum possible score was 20 for all four tasks I, II, III and T.

In each case when a correct response was made the red light came on and the experimenter said 'Good'. If the response was incorrect, the experimenter said 'No', the red light failed to come on, and the trap-door remained closed. The card was returned to the subject who was then required to try again. In the rare cases where a correct response was not obtained after five attempts, the experimenter placed the card in the correct compartment and the next response card was offered.

Note was taken of each response and the precise location of each error. In order to minimize the possibility of rote learning, the positions of the stimuli on the panel were varied for each trial by moving them one place to the right, with the extreme right-hand card becoming the extreme left-hand.

The following scores were available: (1) number of correct first placements per trial out of a possible total of 20; (2) number of errors per trial, with a pre-set limit of five errors per response card (a relatively infrequent occurrence).

The groups followed the experimental design outlined above: training group I received 24 trials (2 per day on alternate days) on task I. Training group II followed the same procedure on task II. Training group III received six trials on task II to induce a set for shape, followed by 18 trials on task III. The control group received no intervening training between test and retest. The mean initial scores and errors for the training tasks are given in Table 1.

After the training period, all four groups were retested on task T to establish the amount of immediate transfer. Thereafter they were all given nine further trials on the same transfer material T to establish whether differences in direct acquisition would relate to earlier differential transfer.

RESULTS

Table 1 gives 'first scores' (i.e. number of placements correct first time) and 'errors' for the pretest and retest on task T. The statistical analyses to be presented are confined to scores, since errors followed the same general pattern, and the scores are the more stringent measure of concept availability.

Since the range of scores on T was considerable, the differential effect of training on low and high initial scorers appeared to be of importance. (Age and place of residence were unrelated to amount of transfer; within the three experimental groups, however, there was a significant inverse relation with I.Q. ($P < 0·025$), as would be expected in view of a significant correlation between initial score and I.Q.) A factorial analysis of variance was computed on the difference scores between pretest and retest contrasting initial scores of 9 out of 20 and above with those of 8 and below. (see Table 2). Since the numbers in the cells are unequal, the method of least squares was used (Winer, 1962). Tukey's method of multiple comparisons of means was used (Ryan, 1959) since it is more stringent than *t*-test comparisons.

Table 2. *Pretest/retest differences for sorting score on transfer task T*

Data summary

		Group			
		II	III	I	Control
Low pretest scorers	*n*	4	5	5	5
	Mean	6·75	5·80	3·80	0·80
High pretest scorers	*n*	5	4	4	4
	Mean	4·30	1·75	0·50	−3·75

Analysis of variance (least squares)

Source	Sum of squares	D.F.	Mean square	*F*-ratio	*P*
High/low	115·99	1	115·99	13·08	< 0·005
Treatments	236·72	3	78·91	8·89	< 0·001
High/low × treatments	5·12	3	1·71	< 1	N.S.
Residual	248·45	28	8·87	—	—

Table 3. *Tukey comparison of means for pretest/retest sorting scores*

	Group			
	II	III	I	Control
Difference score means	5·33	4·00	2·33	−1·22
Group II	—	N.S.	$P < 0·10$	$P < 0·01$
Group III	—	—	N.S.	$P < 0·01$
Group I	—	—	—	$P < 0·05$
Control group	—	—	—	—

The results show that there is a significant tendency for the low initial scorers to transfer more than the high scorers. This is particularly relevant to group I, which occupies a position intermediate between groups II and III and the control group. The positive effect of practice on a simple discrimination task was demonstrated for the low but not for the high scorers, whereas for the groups receiving more complex training both high and low scorers showed transfer gains.

The approximate stability of the control group's mean T scores between pretest and retest shows that the differential increments of the experimental groups resulted from their training conditions (see Table 3). The different trends shown by high and low initial scorers in the control group may be roughly consistent with the sort of regression effect which might be expected in tasks of this kind where each response card carried a one-in-five probability of correct placement by chance. Further comments of the differences between the groups are made later in the article.

The second half of the experiment involved training of all groups, including control subjects on the transfer task, for a further nine trials. Thus, all subjects had a total of eleven trials: a pretest, a retest and nine training trials on this task. Fig. 1 shows the average scores for all four groups at the various stages of the experiment. To test the hypothesis that previous complex training would differentially enhance the level of acquisition of task T, an analysis of variance was computed for trials 2–11. The

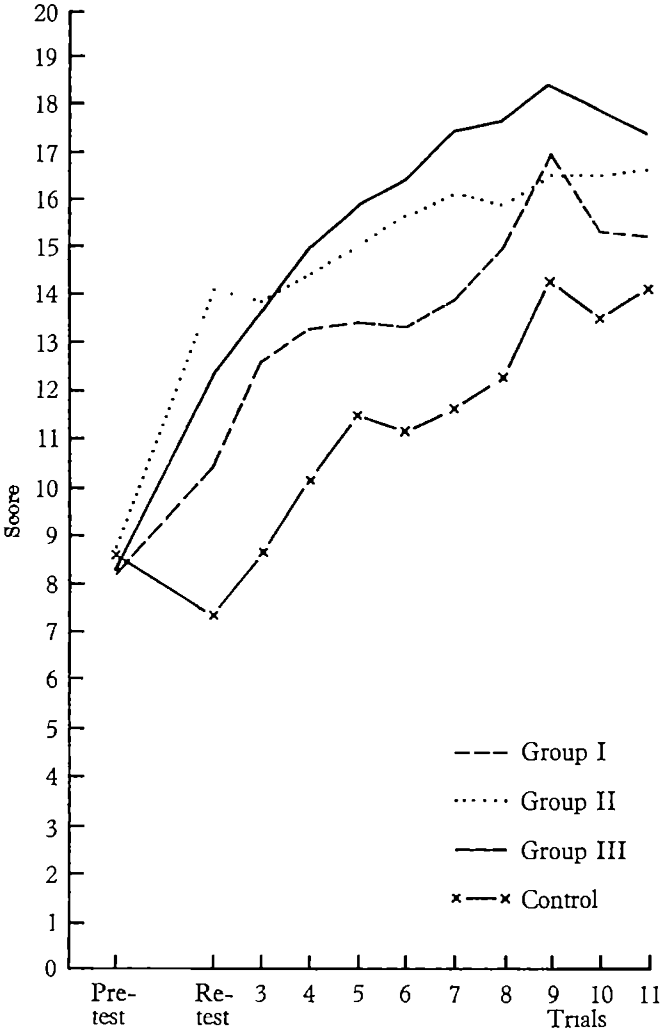


Fig. 1. These graphs show the 'first correct placement' scores of the four groups when sorting twenty photographs of objects belonging to five conceptual categories (human beings, animals, furniture, tableware and clothes), task T. The groups were equated on pretest scores and randomly allotted to control and to three training conditions of different complexity which took place between pretest and retest. Thereafter direct acquisition trials on task T were given. Results indicate the effect of differing complexity of training upon retest, and the maintenance of these differences during later trials.

results are summarized in Table 4. They are consistent with those based on an analysis of pretest–retest differences. It appears that training on a fairly simple matching task (group I) gave these subjects a significant advantage over the members of the control group, and that training on either of the two complex tasks (groups II and III) resulted in an enhanced performance for these subjects over that for group I. A limiting factor in the use of Tukey's method is that Studentized ranges are to the best of our knowledge available at the 0·10, 0·05 and 0·01 levels only.

Table 4. *Analysis of variance and comparison of means for sorting scores on task T, trials* 2–11

Analysis of variance

Source	D.F.	Sum of squares	Mean square	F	P
Between groups	3	1142·37	380·79	34·27	$< 0{\cdot}001$
Between occasions	9	963·99	107·11	9·64	$< 0{\cdot}001$
Interactions	27	88·85	3·29	< 1	N.S.
Residual	310	3444·22	11·11	—	—

Comparison of means

	Group			
	III	II	I	Control
Means...	16·18	15·49	13·98	11·39
Group III	—	N.S.	$P < 0{\cdot}01$	$P < 0{\cdot}01$
II	—	—	$P < 0{\cdot}01$	$P < 0{\cdot}01$
I	—	—	—	$P < 0{\cdot}01$
Control	—	—	—	—

Group III showed no superiority over group II on retest (the difference between them is insignificant), and although they went on to surpass group II during training on task T, the difference between them remained statistically non-significant. Thus, we must conclude (*a*) that the type of task represented by II contained the optimum conditions for producing transfer; or (*b*) that the nature of the complexity dimension superimposed on task II in task III was irrelevant to the transfer situation; or (*c*) that since group III received six trials on task II and eighteen on task III, they had not reached the optimum level of learning on the latter task.

The relative difficulty of the transfer task is underlined by the fact that, even after nine further trials of training on it, the mean scores for the controls was only 13·5, and for training-group I the highest mean score achieved was 16·8. A clustering of those with scores of 17 and above was very noticeable in groups II and III. At trial 11 the distribution of scores throughout the groups was as follows:

	Score	
	17–20	Below 17
Groups II and III	12	5
Group I and control	5	12

(The Fisher exact probability test, two-tailed, gives $P = 0{\cdot}0381$. By this stage one subject had been lost from group III and one from the control group.) It is thus clear that after a total of eleven trials on task T for all the groups, those subjects who were performing at a high level of proficiency were found predominantly in the two groups which had received complex training on the geometric tasks.

Analysis of the location of both errors and first correct placements revealed that gains were similar across all five concepts for each group, the relation between each category remaining approximately the same over training. The most difficult concept among the five was 'clothes', possibly because, of all the objects dealt with, these were the only ones not commonly seen except in association with something else, e.g. a human being or an article of furniture. Acquisition of the other concepts was clearly easier, and differences between these four were small.

DISCUSSION AND CONCLUSIONS

The hypotheses were confirmed: (1) transfer can be demonstrated in relatively naïve subjects across tasks differing considerably in perceptual and conceptual content; (2) transfer is related to the relative difficulty of the task as shown both by the differences across the treatment groups, and also by the differential behaviour of the high and low scorers within the groups, the lower scorers showing a significantly greater amount of transfer than the higher scorers. It appears that transfer from a different task can be sufficient substantially to alter the level of subsequent direct acquisition of a task; this is shown by the clear superiority of groups II and III over group I and the control group during the latter part of the experiment.

The results further suggest that had the training tasks been made more difficult for the initial high scorers, their transfer gains might have been as great as was demonstrated for the low scorers.

In interpreting the results four possibilities have been considered.

(1) That there was a similarity inherent in the stimulus material is obviously not so. No stimulus generalization can have occurred, nor can a conceptual principle have been generalized such as occurred in Judd's experiment.

(2) That the results are at least partly due to 'warm-up' or increased task familiarity: this is thought unlikely on the grounds that the subjects trained on the simple task (group I) achieved a virtually perfect performance on their first trial (mean score 19·5/20).

(3) That the initial scores gave a spuriously low indication of concept availability owing to gross inattention on the part of the subjects: this possibility might suggest itself because of the greater increases occurring among the initial low scorers. It is rejected on the basis of observation of the subjects' behaviour during the experimental sessions, and also because the initial error scores, though high, were not such as to suggest that the subjects were operating largely at random. Furthermore, results from the control groups do not support such a contention.

(4) That the results were due to the experimenters' expectations: this hypothesis is one which is too rarely raised in connexion with human learning experiments, and attention has recently been drawn to it by Bell (1965) writing on methods in developmental psychology. The evidence against this is that group III did not show more transfer on retest than group II, as had been hoped. Furthermore, the greater gains shown by the initial low scorers had not been anticipated, although it accords well with the general postulation of a relationship between task difficulty and transfer.

If we accept Kendler's (1961) definition of concept behaviour as 'the acquisition or utilization or both of a common response to dissimilar stimuli', are we to conclude that the sorting of geometric shapes of varying degrees of complexity *created* in some way more efficient concepts of furniture, tableware, etc? Since the common elements in our tasks were the act of sorting and the act of categorizing, it seems that practice in the act of making common responses to dissimilar stimuli may well have transferred to a widely different task. This tentative conclusion, of course, is based on the assumption that for transfer to occur there must be some similarity between the tasks, and if no such similarity is immediately apparent, the search for identical elements is pursued into the realms of response organization. In fact until a great deal of further work has been done, using young children or other naïve subjects such as imbeciles, it will neither be known how far the idea of width of transfer can be taken, nor indeed whether in certain circumstances the concept of similarity/dissimilarity will be pressed so hard as to cease to be useful.

The formulation proposed by Clarke & Cookson (1962) of increased perceptual and conceptual 'sharpening' is germane to the findings of this experiment. It is probable that attention to the details of the task in hand was controlled by the two more difficult training tasks (II and III) and transferred to the transfer task T, while practice on the simple task (I) produced some increased attentiveness for the initial low scorers, but failed to touch the more sophisticated members of the group.

We believe it unlikely that these findings will apply only to a severely subnormal population, although caution must be exercised in generalizing from this population to others. A further modified experiment on normal nursery-school children has been carried out to test the hypothesis that transfer as a function of task complexity will be demonstrable in any relatively unsophisticated population. The results, so far unpublished, are highly significant and substantially similar to those presented here.

The writers gratefully acknowledge the provision of excellent research facilities by Dr A. Hutchison, Medical Officer of Health, Dr W. Fairlamb and his staff, Kingston-upon-Hull, and by Dr G. W. Waters and his staff at Brandesburton Hospital; helpful advice from Mr J. P. N. Phillips, Mr A. B. Royse and Prof. A. D. B. Clarke; the great assistance of the University Photographer, Mr A. Marshall; the construction of card-sorting apparatus by the Chief Technician, Mr J. Devine, and assistance from his staff.

The investigation was made possible by a generous grant from the Nuffield Foundation.

REFERENCES

BELL, R. Q. (1965). Developmental psychology. *Annu. Rev. Psychol.* **16**, 1–38.

CLARKE, A. D. B. & BLAKEMORE, C. B. (1961). Age and perceptual-motor transfer in imbeciles. *Brit. J. Psychol.* **52**, 125–31.

CLARKE, A. D. B. & COOKSON, M. (1962). Perceptal–motor transfer in imbeciles: a second series of experiments. *Brit. J. Psychol.* **53**, 321–30.

CLARKE, A. D. B. & COOPER, G. M. (1966). Age and perceptual–motor transfer in imbeciles: task complexity as a variable. *Brit. J. Psychol.* **57**, 113–19.

JUDD, C. H. (1908). The relation of special training to general intelligence. *Educ. Rev.* **36**, 28–42.

KENDLER, T. S. (1961). Concept formation. *Annu. Rev. Psychol.* **12**, 447–72.

HENDRICKSON, G. & SCHROEDER, W. H. (1941). Transfer of training in learning to hit a submerged target. *J. educ. Psychol.* **32**, 205–13.

RYAN, T. A. (1959). Multiple comparisons in psychological research. *Psychol. Bull.* **56**, 26–47.

WINER, B. J. (1962). *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

(*Manuscript received* 6 *September* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, pp. 129–135
Printed in Great Britain

CREATIVITY AND INTELLIGENCE: A PARTIAL REPLICATION WITH SCOTTISH CHILDREN OF GETZELS' AND JACKSON'S STUDY

By PARWEEN HASAN AND H. J. BUTCHER
University of Edinburgh

Getzels & Jackson (1962) found that, when two contrasting groups of children were formed, one of high scorers on tests of divergent thinking or 'creativity', the other of high I.Q., (*a*) both groups scored equally on tests of attainment, (*b*) teachers rated the 'high creativity' group rather lower on desirability as pupils. In the whole sample (mean I.Q. 132) from which the groups were drawn correlations between I.Q. and measures of creativity were relatively low. Similar tests were given to 175 Scottish children (mean I.Q. 102). Correlations between I.Q. and measures of creativity were much higher, in spite of a smaller range of ability. When contrasting 'high creativity' and 'high I.Q.' groups, it was found that the 'high I.Q.' group was preferred by teachers, as in Getzels and Jackson's study. Contrary to their findings, the 'high creativity' group was lower in attainment. Results are also presented for children who were high scorers in both 'creativity' and 'intelligence'. Possible reasons for the discrepancies are discussed.

During the last 15 years, there has been a greatly increased interest in the cognitive aspects of 'creativity', which stems to a large extent from the work of J. P. Guilford and his associates. In particular, his distinction between convergent and divergent thinking and the construction of new kinds of test to measure the latter have led many psychologists to believe more strongly that important kinds of ability have not been adequately assessed by conventional intelligence tests. One of the best-known studies and one that makes some of the boldest claims about the possibility and usefulness of distinguishing 'creative' from 'intelligent' children is that of Getzels & Jackson (1962). Their findings are important if valid, but were obtained from a special and highly selected group of American children, and by methods that have not escaped criticism.

Getzels & Jackson (1962) studied 449 boys and girls, aged from 12 to 17 years, who were pupils at a private school in the Chicago area, and whose average level of ability was very high (mean I.Q. 132; S.D. 15). The general aim was to try to identify from among these pupils 'two groups of students each exhibiting as its salient characteristic a different type of cognitive excellence—the one high in intelligence but not concomitantly high in creativity, the other high in creativity but not concomitantly high in intelligence', and to study and compare these groups in terms of their behaviour in school, degree of attainment, popularity, values and other characteristics. The 'high intelligence' group were those children who were in the top 20 % in terms of I.Q. but not in their scores on a battery of tests of creativity; the 'high creativity' group were in the top 20 % in creativity score but not in I.Q. The creativity battery was composed of five tests, some of which were originally devised by Guilford and by Cattell. These were called Word Association, Uses for Things, Hidden Shapes, Fables, Make-up Problems.

Differences were found between the 'high creativity' and 'high I.Q.' groups. Most striking perhaps was the finding that the high creativity group equalled the high

I.Q. group in scholastic achievement (as assessed by a composite score on several standardized tests) in spite of having an average I.Q. which was 23 points lower. Getzels and Jackson also attempted to find out whether this unexpectedly high degree of scholastic achievement on the part of the high creativity group could be ascribed to higher strength of motivation. They found, however, no difference between the two groups on McClelland's 'need for achievement' measure, and they concluded that the explanation lay not in any motivational difference, but in the predictive limitations of the conventional intelligence test.

Another finding was that teachers appeared to approve more strongly of the high I.Q. group than of the high creativity group. The average rating for 'desirability as a student' was slightly higher for the former group, even though it was made clear in the rating instructions that 'brightness' was not to be taken specifically into account. Getzels and Jackson commented that the reverse should have been true, since the high I.Q. group were only doing scholastically what was expected of them, whereas the high creativity group (as shown by their achievement) were doing relatively more than might have been expected of them.

De Mille & Merrifield (1962) criticized Getzels' and Jackson's study as being ill-designed and inadequately reported. They were particularly critical of the failure to provide any information about children who were high scorers in tests both of divergent and of convergent thinking. Burt (1962), Marsh (1963) and Vernon (1964) have all maintained the view that much of what has been described as 'creativity' may be re-analysed and shown to be ascribable to a factor of general intelligence. Dunnette (1964) pointed out that some differences claimed as important were not statistically significant.

Torrance (1962) repeated much of Getzels' and Jackson's work in several relatively non-selective schools with varying results. In about half of these replications the results were doubtful or negative, but Getzels' and Jackson's findings were supported in some cases. Yamamoto (1965) has also summarized more recent positive evidence. Edwards & Tyler (1965), on the other hand, working with a sample in a non-selective American junior high school, found almost entirely negative results and concluded that Getzels' and Jackson's findings about the relation of creativity, intelligence and achievement were not widely generalizable. No repetitions of these experiments in Britain appear to have been reported.

The aim of the present investigation was therefore to throw some light on the generality of Getzels' and Jackson's findings, particularly those about relations between creativity, intelligence and attainment. Some slight modifications were made in the tests used and in the methods of reporting results. The main difference, however, was in the attempt to apply their methods to a group of children unselected for ability.

METHOD

Plan of the investigation

The investigation was a partial replication of Getzels' and Jackson's work with a number of modifications.

(1) It was limited to considering relations between performance on (*a*) a test of verbal reasoning or 'intelligence'; (*b*) tests of divergent thinking or 'creativity'; (*c*) English and arithmetic attainment tests; and (*d*) teachers' ratings on desirability as a pupil. Personality factors and family background could not be investigated.

(2) Additional tests of creativity besides those used by Getzels and Jackson were included. One of their five tests had to be omitted.

(3) In addition to forming contrasting groups of children with high verbal-reasoning quotient (VRQ) and with high scores on tests of creativity respectively, a group which was high on *both* VRQ *and* creativity has also been included.

(4) The Scottish sample of pupils differed from Getzels' and Jackson's by being (*a*) more homogeneous in age and (*b*) more typical in level of ability of the general population of schoolchildren.

Tests and other measures used

Testing was carried out at a Scottish local authority comprehensive school, with 175 pupils (100 boys, 75 girls) in their second year of secondary education. The following procedures were used.

Verbal-reasoning quotients were each child's mean score on two verbal reasoning tests administered by the local education authority for the purpose of secondary-school placement.

Arithmetic and English quotients (AQ and EQ) were obtained from Moray House English and Arithmetic attainment tests given during the course of the experiment.

Teachers' ratings on 'desirability as a pupil' were obtained by asking teachers to rate all subjects on a five-point scale with values of 15, 12, 9, 6, 3 for each category.

The creativity score was the aggregate of scores obtained from ten tests of creativity. (i) Word-Association (Getzels & Jackson, 1962). (ii) Uses for Things (Getzels & Jackson, 1962). (iii) Completing Fables (Getzels & Jackson, 1962). (iv) Make up Problems: modified from Getzels & Jackson (1962). (v) Word-Fluency (from Thurstone's P.M.A. battery). (v) Expressional Fluency: based on Guilford's (1956) model. (vii) Remote Associations Tests (2): derived from Mednick's (1962) model, but with two instead of three words in each item, (viii) Remote Associations Test (3): as above, but with three words in each item. (ix) Drawing: subjects were given various geometrical and other shapes (triangle, teardrop, oval, etc.) drawn on paper; they were required to make as many objects as possible by joining these shapes on tracing paper. (x) Circles (Torrance, 1962).

Methods of analysis

Scores were treated in two main ways in order to make the results comparable with those of Getzels and Jackson. Intercorrelations were computed for the whole sample (*a*) among the creativity tests, (*b*) between creativity and VRQ, (*c*) between creativity and attainment and (*d*) between VRQ and attainment. Two contrasting experimental groups were selected in exactly the same way as by Getzels and Jackson; in addition, a third experimental group was formed of those children who scored high both in creativity and intelligence. Scores of the three experimental groups on all four kinds of variable were compared.

RESULTS

Relations between creativity and intelligence

The intercorrelations of the ten 'creativity' measures, the correlations of each with the aggregate 'creativity' score derived from them, and the correlations of all eleven 'creativity' variables with the Moray House Verbal Reasoning score are shown in Table 1. Every creativity measure correlates positively and significantly with VRQ and these correlations are generally rather higher than the intercorrelations between the measures of creativity. Indeed, the correlations of individual tests of creativity with VRQ are in most cases not very much lower than their correlations with the creativity aggregate, although the latter are part–whole correlations.

It is not possible to make a complete comparison between Table 1 and the results of Getzels and Jackson, first because not all the measures of creativity were the same in both studies, and secondly because almost all Getzels' and Jackson's correlations are reported for boys and girls separately and not for the entire group. But three of the tests of creativity were the same in both studies, and the results for them were also analysed for boys and girls separately (cf. Table 2).

Table 1. *Intercorrelations of the ten tests of creativity, creativity aggregate and VRQ*

Boys and girls, $n = 175$. Coefficients in italics do not differ significantly from zero ($P > 0·05$).

	1	2	3	4	5	6	7	8	9	10	11	1
1 VRQ	—	0·741	0·518	0·682	0·438	0·213	0·185	0·345	0·544	0·260	0·240	0·7
2 Word Association	—	—	0·481	0·632	0·453	0·178	0·188	0·318	0·492	0·302	0·347	0·7
3 Uses	—	—	—	0·512	0·381	*0·145*	*0·062*	0·281	0·403	0·187	0·413	0·6
4 Fables	—	—	—	—	0·416	*0·131*	0·163	0·358	0·448	*0·164*	0·253	0·7
5 Problems	—	—	—	—	—	*0·101*	0·182	0·186	0·301	0·209	*0·070*	0·6
6 Word Fluency	—	—	—	—	—	—	0·188	*0·027*	*0·134*	*0·017*	*0·041*	0·2
7 Expressional Fluency	—	—	—	—	—	—	—	*0·005*	0·153	*0·077*	0·255	0·4
8 Remote Associates (2)	—	—	—	—	—	—	—	—	0·583	0·241	*0·072*	0·5
9 Remote Associates (3)	—	—	—	—	—	—	—	—	—	0·170	0·269	0·6
10 Drawing	—	—	—	—	—	—	—	—	—	—	*0·124*	0·3
11 Circles	—	—	—	—	—	—	—	—	—	—	—	0·4
12 Creativity Aggregate	—		—	—	—	—	—	—	—	—	—	—

Table 2. *Correlations of three tests of creativity with intelligence* (*VRQ*)

The coefficients in brackets are the corresponding correlations with intelligence from Getzels & Jackson (1962)

	Boys		Girls	
n . . .	100		75	
Word Association	0·772	(0·378)	0·598	(0·371)
Uses	0·551	(0·186)	0·317	(0·147)
Fables	0·726	(0·131)	0·468	(0·115)

The general impression from Table 1 of much more overlap between creativity and intelligence than was found by Getzels and Jackson is confirmed by the direct comparisons. It might be thought that the much higher correlations were a purely statistical effect due to the relatively unselected nature of the Scottish sample, but this is not so. Although the Scottish sample is more typical of the general run of children in level of ability, the actual spread of ability is less than in the American sample.

Relations of creativity and intelligence to attainment

This question was examined in two ways. First, correlations were calculated for the entire Scottish sample (cf. Table 3), and secondly the attainment scores of contrasting groups were compared, as in the American study (cf. Table 4).

Table 3. *Correlations of creativity and intelligence* (*VRQ*) *with measures of attainment*

Boys and girls, $n = 175$.

	Attainment	
	Arithmetic	English
Creativity aggregate	0·615	0·758
VRQ	0·734	0·845

Table 4. *Mean scores and standard deviations of the whole sample and of the three experimental groups on VRQ, AQ, EQ and Creativity Aggregate*

	Whole sample ($n = 175$)		High creativity group ($n = 17$)		High VRQ group ($n = 17$)		Group high on both creativity and VRQ ($n = 24$)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
VRQ	102	12	106	5	115*	7	117*	5
Arithmetic	108	13	113†	13	123‡	8	119‡†	14
English	102	14	108	7	114	8	119	12
Creativity Aggregate	140	40	188	16	152	13	190	18

*, †, ‡ Mean scores marked with the same sign *do not* differ significantly ($P > 0{\cdot}05$). Since the sample includes the experimental groups, *t*-tests were applied only to differences between experimental groups.

Table 4 shows that the 'high creativity' group obtained significantly lower scores on both measures of attainment than did the 'high intelligence' group. The children in the 'high creativity *and* high intelligence' group scored highest of all in English, and their average was not significantly lower than that of the 'high intelligence' group in arithmetic. Table 4 also confirms the high degree of overlap between 'creativity' and 'intelligence' in the Scottish sample implied by Table 1. As can be seen, it was not possible to form a 'high creativity' group which had the highest scores on the aggregate of the creativity tests and which, at the same time, excluded children with the highest VRQ.

Teachers' ratings of 'creative' and 'intelligent' children on desirability as pupils

The mean ratings for both the 'high VRQ' and the 'high on creativity *and* VRQ' groups are significantly more favourable than the mean for the 'high creativity' group (cf. Table 5). Getzels' and Jackson's findings of relatively low approval for 'creative' children thus receives some confirmation. Two points, however, need to be borne in mind. It would seem not to be the higher 'creativity' as such, but to be the lower intelligence that affected the teachers' ratings, since the group high on both (and highest on 'creativity', cf. Table 4) received the highest ratings of all. Secondly, Getzels and Jackson contrasted the relative lack of approval received by the 'creative' children with their unexpectedly high attainment, suggesting that it was therefore particularly undeserved. No such conclusion can be drawn from our results.

Table 5. *Mean scores and standard deviations of teachers' ratings of 'desirability as a pupil' for the whole sample and for the three experimental groups*

		Whole sample ($n = 175$)	High creativity group ($n = 17$)	High VRQ group ($n = 17$)	Group high on both creativity and VRQ ($n = 24$)
Desirability as a pupil	Mean	10·5	10·5	12·3	12·6
	S.D.	2·9	2·6	2·7	2·8

DISCUSSION AND CONCLUSIONS

The results presented must strongly suggest that some of Getzels' and Jackson's findings are not generally applicable. Some further evidence and a more detailed account of the experiment described here can be found in Hasan (1965). In some atypical groups of children it may be as easily possible to distinguish 'creative' from 'intelligent' children as has been claimed, but it appears that in a relatively unselected group measures of convergent and of divergent thinking very largely overlap, to such an extent indeed as to lend support to Burt's (1962) remark that the latter would 'form very satisfactory additions to any ordinary battery for testing the general factor of intelligence'. Nor is this merely, as Hudson (1965) calls it, 'a rather academic point'. High hopes have been raised that it might be possible to distinguish by means of open-ended tests of divergent thinking a set of able children whose abilities have hitherto been concealed or minimized by the use of conventional tests. Our results seem to indicate that this will not easily be possible in an unselected group of children by means of the open-ended tests at present available.

The two main points on which our results differ from those of Getzels and Jackson are first the degree of overlap just referred to, and secondly our failure to find that children who score high on tests of divergent thinking (in spite of relatively low conventional I.Q.s) are thereby likely to show up better than would otherwise have been expected on tests of attainment. Admittedly our criteria of attainment were narrow, standardized and therefore perhaps particularly likely to correlate highly with tests of intelligence or verbal reasoning. But the criteria of achievement used by Getzels and Jackson were also standardized tests, and the large difference in results is unlikely to be ascribable entirely to this cause.

One possible explanation of the discrepancy is the 'threshold' theory (Mackinnon, 1962), according to which 'above a certain required minimum level of intelligence, which varies from field to field and in some cases may be surprisingly low, being more intelligent does not guarantee a corresponding increase in creativeness'. Barron (1963) has suggested a threshold of about I.Q. 120, above which intelligence ceases to be so relevant to many forms of achievement. If this theory is valid, Getzels and Jackson's findings might have received more support if their study had been repeated with another group of particularly able children.

Cultural differences between Chicago and Edinburgh may be relevant, but the varying success of the attempts to establish Getzels' and Jackson's claims in other American studies suggests that it cannot be the whole explanation. A more specific hypothesis about differences between kinds of school, which may interact with national differences, has recently been put forward by Torrance (personal communication) as possibly accounting for the discrepant results; he plausibly suggests, and claims some support for this view from his own experiments, that the success of predominantly divergent thinkers is directly related to the degree of freedom and permissiveness and the lack of authoritarian discipline within a school.

These two theories, one of a threshold effect, and the other of an effect of school atmosphere on divergent thinking and 'creativity', need to be tested.

REFERENCES

BARRON, F. (1963). *Creativity and Psychological Health.* Princeton, N.J.: Van Nostrand.

BURT, C. (1962). Critical notice of *Creativity and Intelligence* by Getzels and Jackson. *Brit. J. educ. Psychol.* **32**, 292–8.

DE MILLE, R. & MERRIFIELD, P. R. (1962). Review of *Creativity and Intelligence* by Getzels and Jackson. *Educ. psychol. Measmt.* **22**, 803–8.

DUNNETTE, M. D. (1964). Critics of psychological tests: basic assumptions: how good? *Psychol. in the Schools*, **1**, 63–9.

EDWARDS, M. P. & TYLER, L. E. (1965). Intelligence, creativity and achievement in a non-selective public junior high school. *J. educ. Psychol.* **56**, 96–9.

GETZELS, J. W. & JACKSON, P. W. (1962). *Creativity and Intelligence.* New York: Wiley.

GUILFORD, J. P. (1956). The structure of intellect. *Psychol. Bull.* **53**, 267–93.

HASAN, P. (1965). Creativity and Intelligence. Unpublished B.Ed. thesis, University of Edinburgh Library.

HUDSON, L. (1965). Intelligence: convergent and divergent. *Penguin Science Survey* 1965: *B.* Harmondsworth: Penguin Books.

MACKINNON, D. W. (1962). The nature and nurture of creative talent. *Amer. Psychologist* **17**, 484–95.

MARSH, R. W. (1964). A statistical re-analysis of Getzels' and Jackson's data. *Brit. J. educ. Psychol.* **34**, 91–3.

MEDNICK, S. A. (1962). The associative basis of the creative process. *Psychol. Rev.* **69**, 220–32.

TORRANCE, E. P. (1962). *Guiding Creative Talent.* Englewood Cliffs, N.J.: Prentice Hall.

VERNON, P. E. (1964). Creativity and intelligence. *Educ. Res.* **6**, 163–9.

YAMAMOTO, K. (1965). Multiple achievement battery and repeated measurements: a postscript to three studies on creative thinking. *Psychol. Rep.* **16**, 367–75.

(*Manuscript received* 11 *August* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 137–153
Printed in Great Britain

THE GENETIC DETERMINATION OF DIFFERENCES IN INTELLIGENCE: A STUDY OF MONOZYGOTIC TWINS REARED TOGETHER AND APART

BY CYRIL BURT
University College London

In the course of a series of investigations into various aspects of mental inheritance an intensive study has been made of so-called 'identical' twins. The cases examined fall into two main groups: first, those reared together in their parents' homes; secondly, those separated in early infancy, and brought up apart. With the latter, despite wide differences in environmental conditions, the correlations for intelligence, unlike those for school attainments, prove to be surprisingly high. It is argued that this implies that 'intelligence', when adequately assessed, is largely dependent on genetic constitution. Supplementary correlations from other groups (dizygotic twins, siblings reared together and apart, first and second cousins, and other related pairs) confirm this conclusion.

I. THE COMPLEXITY OF THE PROBLEM

Individual psychology and psychogenetics might themselves be fairly described as fraternal twins: they were begotten by the same father, Sir Francis Galton, and they emerged almost simultaneously as vigorous offspring of the mother science. Unfortunately they became parted soon after birth. Consequently, as Professor Darlington has remarked in a recent paper (this *Journal*, 1963, p. 293), 'despite their common origin, psychology and genetics, whose business it is to explain behaviour, have failed to face their task together'. One reason suggested is that general psychology has so far been unable to supply the genetic psychologist with any clear conception of what is commonly called the mind, or of its structure and development, such as might serve as a working basis for his researches. He has therefore to shoulder the preliminary task of determining for himself what particular traits or tendencies he shall select for observation and assessment.

In the case of the human species the variations with which the student of individual psychology is commonly concerned have proved to be extremely complex. In the first place, as Galton pointed out, a person's success in almost every walk of life depends on at least three distinguishable types of mental quality—cognitive, affective and conative: 'he must possess the requisite abilities; he must respond with eagerness and zeal; and he must sustain the necessary effort'. Secondly, statistical studies based on factorial techniques have demonstrated pretty clearly that we can no longer be content with the traditional notion of a motley assortment of cognitive 'faculties' or 'primary abilities', such as Thorndike at one time postulated and many educational and psychological writers still tacitly assume, nor yet with the over-simplified hypothesis of a single 'general factor' which Spearman proposed to substitute. The evidence—biological and neurological as well as purely statistical—suggests that the brain, or rather the central nervous system, is an organized hierarchy comprising both a 'general cognitive factor' and a number of more specialized 'group factors' of varying extent or breadth; but of these the 'general factor' is by far the most important, especially during early years. Thirdly, a vast mass of evidence, which I have summarized in previous papers, suggests that differences in this factor are largely the effect of the individual's genetic constitution—the product of the particular set of genes which each pair of parents transmits to their offspring—and that the mode of transmission conforms to the general principles worked out by Mendel and his followers for unifactorial and multifactorial inheritance (Burt, 1912, 1958, and refs.).

This threefold distinction—between cognitive characteristics and motivational, between general characteristics and the more specialized, and between genetic characteristics and those that are acquired—leads to the concept of an 'innate, general, cognitive factor', each of the adjectives being defined in a somewhat technical sense. The problem for the psychologist therefore has been to decide, by means of appropriately planned research, whether such a factor really exists, and if so, what is its practical significance. To designate this concept it is more convenient to use one word instead of four, and Spencer's term 'intelligence' has been adopted (perhaps a little unwisely*) as a popular label. Differences in this hypothetical ability cannot be directly measured. We can, however, systematically observe relevant aspects of the child's behaviour and record his performances in standardized tests; and in this way we can usually arrive at a reasonably reliable and valid estimate of his 'intelligence' in the sense defined. But it is clear that what we thus observe or calculate is a somewhat artificial abstraction, and, developmentally at any rate, decidedly remote from the aggregate of biochemical tendencies transmitted at the time of sexual reproduction.

Of late, however, an increasing number of British psychologists and educationists have vigorously challenged, not so much the bare fact of mental inheritance, but the idea that it has any appreciable importance as compared with environmental influences. Most of their criticisms rest, not on any fresh evidence or new researches of their own, but chiefly on armchair arguments from general principles. There are two obvious ways in which the questions thus raised can be met and dealt with.

(1) First, we can compare the performances of children who *differ in their presumable inheritance*, but have been brought up in much the same environment—children, for example, who have been removed soon after birth to an orphanage or other institution. With the present problem largely in view, I and my colleagues under the London County Council carried out a series of investigations on such cases (Burt, 1943). Two main results emerged. (i) In spite of the uniformity of the environment, the individual variation in intelligence was at least as great as that of a random sample of children living in their homes. (ii) These variations showed a fairly close correlation (about 0·5), with the variations in the intelligence of their parents. Some of the most striking instances were illegitimate children of high ability; often the father (as the records showed) had been just a casual acquaintance of the mother, well above her own social and intellectual status, who had taken no further interest in the child, and whom the child himself had never even seen. In such a case it would be out of the question to attribute the child's exceptional intelligence to the cultural opportunities of his home, since his only home had been the institution to which he had been sent.

---

* How unwisely is shown by the use of the term in the recent correspondence in the *Listener* (24 June to 17 July 1965), which followed articles on mental inheritance by Dr Liam Hudson and Dr Hammerton. Almost everybody treated the word 'intelligence', not as a technical term, but as a word of popular speech to be interpreted according to the fancy of the writer.

(2) Secondly, however, we may compare the performance of children who have the *same genetic constitution*, but have been brought up in widely different environments. Galton (1883) was apparently the first to recognize that the occurrence in man of so-called 'identical' twins 'makes it possible to show, more clearly than in other animals, the relative influence of differences in genotype and in environment respectively' (Darlington & Mather, 1949, p. 349). 'Monozygotic' or 'uniovular' twins are derived from the splitting of a single fertilized egg or zygote. Since they are developed from the same cell-nucleus, it follows that (barring accidents to the chromosomes, which are by no means unknown) they must have the same outfit of genes. However, the cytoplasm of the cell, as well as its nucleus, can occasionally influence heredity; and, unlike the splitting of the nucleus, the splitting of the cytoplasm could introduce a slight genetic difference (Spiegelman, 1946): hence the description 'identical' may be somewhat misleading. In the paper just quoted Darlington noted 'what an extraordinary success this [line of research] has had as a means of comparing the effects of a difference in heredity and in environment'; but he went on to warn the psychologist against the hasty inferences and simplifications often made in interpreting such data, not so much by Galton himself as by some of his over-enthusiastic successors.

In most of the investigations in which 'identical' twins have been distinguished from 'fraternal' the two members of each pair have nearly always been brought up together, usually in their own homes. Now it is well known that identical twins tend to keep together far more than fraternal twins, particularly since about half the fraternal twins are of different sex. The environmentalist therefore naturally argues that the higher correlation for intelligence found in the case of identical twins can be fully explained by the greater similarity in their life-histories. To meet this criticism my co-workers and I decided to make a special study of cases in which the members of such pairs had been brought up in entirely different environmental conditions.

In our earliest survey of children in L.C.C. schools and institutions we reported a number of case-studies in which 'identical' twins had been separated during the first few months of life (Burt, 1921 and refs.); and over the years the number has steadily increased. Other writers, chiefly geneticists, have described isolated instances in which the members of a single pair were reared apart: of these the earliest is perhaps the most instructive (Popenoe, 1922; cf. Muller, 1925). Three group-studies, in addition to our own, have since been published (Newman, Freeman & Holzinger, 1935; Juel-Nielsen & Mogensen, 1957; Shields, 1962); and the total number thus investigated now amounts to well over hundred. Strange to say, however, apart from our own, nearly all the inquiries have been concerned, not so much with children, but with adults; and for adults of course it is far more difficult to secure complete or accurate data.

During the last 10 years or so the inferences drawn from these and similar studies have been repeatedly questioned by several writers, both in this *Journal* and elsewhere—notably by Halsey (1959), Stott (1956), Lewis (1957), Maddox (1957) and Woolf (1952). The most recent and the most outspoken is McLeish (1963). He cites and criticizes the work both of Newman and myself. The correlations obtained, so he tells his readers, show that 'identical twins reared apart resemble each other on intelligence tests about as much as do fraternal twins reared together'; (this is by no means accurate as a glance at the figures will show; see Table 2). At all events, he says, 'the best studies' (which these are he does not specify) 'render the initial assumptions of the tester obsolete or highly questionable'; and the idea that 'identical twins are endowed with the same genes' is itself 'an assumption open to question' (on what grounds we are not told).

On the other hand, Shields (1962) has recently reached conclusions which are much the same as my own. In his review of 'previous reports' he refers to a wartime paper of mine on 'Ability and income', in which I quoted data from our first London survey (Burt, 1921), and included correlations between children related in various degrees (including a small group of monozygotic twins reared apart) whose after histories had been followed up for 15–20 years (Burt, 1943). He regrets that I have 'given no other information about the twins brought up apart'. The paper which he cites, however, was concerned primarily with the 'influence of innate ability and parental income on entrance to universities', and the mention of twins was merely incidental. For further information I expressly referred to previous L.C.C. Reports, and explained that my own research students were still 'working on data obtained for twins up to the outbreak of the war'. After the war a fuller account was printed in the same journal (Burt, 1955), and the statistical evidence set out in some detail by my co-workers and myself in one of the more technical periodicals (Burt & Howard, 1957; Conway, 1958). Both the earlier and the later publications, however, seem to have escaped Dr Shields's attention.

Meanwhile, largely as a result of these various discussions, further cases of separated twins have been brought to our notice, and more information has now been obtained for the earlier cases from the follow-up inquiries. The main purpose of the present paper therefore will be to bring together the evidence now available both from our studies and those of more recent investigators, and at the same time to answer the questions and criticisms raised by Dr Shields, Dr McLeish, and other writers.

II. DATA ON TWINS REARED APART

The tests employed have been fully described elsewhere (Burt, 1921, 1933). Except for three children living in rural areas who had to be visited during the holidays, all the tests were applied in school. They consisted of (i) a group test of intelligence containing both non-verbal and verbal items, (ii) an individual test (the London Revision of the Terman–Binet Scale) used primarily for standardization, and for doubtful cases (iii) a set of performance tests, based on the Pintner–Paterson tests and standardized by Miss Gaw (1925). The test-results, which generally covered other children in the school as well, were submitted to the teachers for comment or criticism; and, wherever any question arose, the child was re-examined. It was not practicable for the same person to test every child. I was helped by three principal assistants, and in a few cases by research-students, all of whom had been trained by me personally. The methods and standards therefore remained much the same throughout the inquiry. If any divergence occured, it would tend to lower rather than to raise the correlations. The reliability of the group test of intelligence was 0·97; of the Stanford–Binet 0·95; of the performance tests 0·87. For school attainments we used the group tests constructed and standardized for London children

(Burt, 1921). Assessments of eye-colour, head-length, and head-breadth were obtained for all twins, but only random samples (fifty of each sex) in the case of the ordinary siblings: measurements for height and weight were corrected for sex and age.* The number of children in each of the categories to be compared is shown at the top of each column in Table 2.

Most writers (e.g. Shields, 1962, pp. vi, 9) apparently suppose that 'monozygotic twins who have been separated from early childhood are of great rarity'. This we believe to be founded on a misconception. At the time of our main survey the number of twins among children born in London (excluding those who were not British) amounted to 1·4%; and of these nearly a quarter must have been monozygotic. This follows because among twins generally about 38% are of unlike sex; and among dizygotic twins the number of pairs of like sex must be approximately equal to the number of unlike sex. Monozygotic pairs are always of the same sex. Hence about $100-2\times 38 = 24\%$ of all twins must be monozygotic. Twins brought up together usually attend the same or neighbouring schools; and, even after full allowance has been made for the higher mortality of twins during pre-school years, the proportion discoverable by school visiting alone is far less than would be expected from the number of those born. What happens is fairly clear. Many mothers are unable or unwilling to rear two children at the same time; but they are generally reluctant for it to be known that they have arranged for one of the children to be removed at or soon after birth. Since the actual placements are often carried out by the local authority or by some public body, a psychologist or social worker who is also a member of the staff can usually obtain full particulars for a large number of such cases. Hitherto most of the published researches have been undertaken by outside investigators who have no access to these confidential records. May we therefore urge that other educational psychologists, who have the advantage of being on the staff of a local authority, should conduct similar inquiries along much the same lines?

From Table 2 it will be seen that the number of monozygotic pairs we have studied now amounts to 148, of whom fifty-three have been reared apart. Of the ninety-five pairs who were reared together, the majority attended London schools; thirty-seven came from areas outside London: nearly half of these were encountered during an investigation I was asked to carry out for the Birmingham Education Authority; several were discovered in the Warwickshire area, where my family lived; and the rest were cases to which our attention was drawn by colleagues or correspondents. In each of the fifty-three pairs reared apart one child at least was a Londoner in all but eight cases. All had been separated either at birth or during their first six months of life. (In our initial report we included three who were separated considerably later; but they were subsequently omitted).

To determine 'zygosity', i.e. to distinguish 'identical' from so-called 'fraternal' twins, no one criterion is sufficient. A difference of sex is of course decisive; such a pair cannot be monozygotic. With the younger children, particularly those born in the area in which the investigator was working, we were frequently able to secure detailed records of the mother's pregnancy and birth; and it is generally agreed that twins reported as born in a single chorion are monozygotic. When twins brought up together can be seen side by side, the impressionistic judgement of an expert observer is likely to be correct in nineteen cases out of twenty; but a few dizygotic twins are remarkably alike. Height, weight, and right- or left-handedness are unreliable; hair-colour, eye-colour, fingerprints and palm prints are more helpful. In doubtful cases the most valuable check is provided by an investigation of blood groups and serum groups. This had not been introduced when we started our inquiries; and, though in half-a-dozen of our later cases where slight doubts existed, it was adopted as an extra precaution, we were unable to carry it out as a routine procedure. We think it highly unlikely that any misclassifications have been made; but, if they have, their effect would be to reduce the differences between the correlations for monozygotic and dizygotic pairs. (See Essen-Möller (1941) and Penrose & Smith (1955) for a discussion of the peculiar methodology of the problem.)

---

* In the initial survey a few of the children living outside London were originally tested by the local teacher or school doctor; but these have all been since re-tested by Miss Conway. I should like once again to express my indebtedness to all who have assisted in this way, particularly to my former assistant Miss V. G. Pelling (who helped with the earlier surveys until her untimely death), to Miss Molteno, Mr Lewis, Miss Howard, and Miss Conway, as well as the various teachers and school doctors who frequently supplied physical and medical data.

One argument often advanced by our critics is that, although our separated pairs may have had a different environment from the day of birth or shortly afterwards, they must have had 'the same pre-natal environment, equally favourable in some cases, equally unfavourable in others'. But this, if correct, would apply to dizygotic twins almost as strongly as to monozygotic twins: the mere fact that the latter are enclosed in the same membrane, as well as in the same uterus, could not of itself account for the far higher correlations they provide. However, the assumptions on which the argument rests are not borne out by the actual facts. To begin with, the position of the foetuses within the uterus, and the time (and often the mode) of delivery will be different. Moreover, the immediate cause of the twinning process appears in most instances to be some temporary setback at the outset of gestation; and this may affect the different foetuses in different degrees. Before splitting the two portions of the developing embryo generally develop at different rates. When the splitting occurs at a fairly late stage, the twin derived from the less developed half tends to be both smaller and weaker. If the splitting occurs after the embryo has begun to assume bilateral symmetry, then, in certain anatomical details, one twin may even be the mirror image of the other. In conjoined or 'Siamese' twins, where the splitting occurs so late as to remain incomplete, reversals seem the rule rather than the exception. Among our own cases we noted that mirror reversals (opposite handedness, reversed patterns in finger-prints and palm-prints, contrary directions in the crown whorl of the hair, etc.) were nearly four times as frequent with monozygotic pairs as with dizygotic. It was among pairs in which, on these and other grounds, we inferred late splitting that we discovered the largest discrepancies in the estimates for intelligence; as a rule, the child delivered second turned out to be the feebler of the two both in mind and in physique: and, as gynaecological records show, it is the one most frequently still-born.

The biggest differences of all were found among low-grade mentally defective twins in L.C.C. instutitions. Here, out of twenty-six cases in which deficiency occurred in a monozygotic twin, there were twenty-one in which the other twin was *not* mentally deficient. These were likewise the cases in which physical differences were most conspicuous. When the deficiency was of a high-grade type, unaccompanied by physical defect, both twins were commonly affected to much the same degree; and among those pairs in which the physical defect occurred in one twin only, the correlation between intelligence and the extent of the physical damage was decidedly low. We therefore concluded that intra-uterine conditions are, if anything, more likely to diminish the resemblance between twins than to increase it, and that, apart from the rare cases of marked pathological deficiency (not included in our present series), their effect on mental capacity is comparatively slight. (For fuller details see Burt & Howard (1956, pp. 123 f.) and references.)

In the records for our cases various reasons were given for the separation of the twins. In nine cases the mother had died either in childbirth or shortly afterwards; in the others the parents generally explained that they had felt unable to bring up both the children 'in a proper way'. In twelve the ground alleged was the mother's poor health; in six she was unmarried; in the remainder the chief or only ground was economic: e.g. the father was dead, out of work, or weak in health, and the mother was the main wage-earner; often, it was said, the family was already too large for the resources of the home; in the better classes the parents frequently declared that they could not afford the cost of educating both children in the way they deserved.

Since our object was to compare twins brought up in different environments with those brought up in similar circumstances, we included in our group of separated twins no cases in which both had been brought up by a relative, except for five in which one relative lived in a town and the other in the country. For our fifty-three separated cases the occupational categories of the parents and foster parents are shown in Table 1. The classification of occupations is the same as that adopted in my previous reports (for details see Carr-Saunders & Jones, 1937, pp. 55f.). The descriptions on the left indicate the occupational category of the children's own parents in cases where one child was brought up by them (the vast majority); in the few remaining cases that of the relative who adopted one of them, or the better type of foster-parent. These figures should dispose of one of the commonest explanations advanced by thoroughgoing environmentalists—namely, that the high correlations for the separated twins is due to the way the foster-parents were chosen.

Table 1. *Occupational categories of parents and foster parents*

		Foster parents							
	Parents	*I*	*II*	*III*	*IV*	*V*	*VI*	Residential institution	Total
I	Higher professional, etc.	0	0	0	1	2	0	0	3
II	Lower professional, etc.	0	0	1	0	4	2	2	9
III	Clerical, etc.	0	1	1	1	3	1	0	7
IV	Skilled	1	1	0	0	2	1	0	5
V	Semi-skilled	2	0	2	1	5	2	3	15
VI	Unskilled	1	1	0	0	7	3	2	14
	Total	4	3	4	3	23	9	7	53

Halsey (1959), for example, criticizes the inferences which my co-workers and I originally drew on two main grounds. First of all, he believes, the practice of 'selective placement' would suffice to account for the similarities found between twins who have been reared apart—an explanation also put forward by Hudson (1965).

Secondly, it is said, our cases 'do not represent the full range of the social and cultural scale'; hence the effects of environmental differences have not been allowed a fair opportunity to reveal themselves. Both criticisms are apparently derived from the notion that (as another critic has put it) 'official surveys confined to pupils in the ordinary elementary schools of the Council would not at that date have included children from the professional or well-to-do classes, nor those who are mentally deficient': with a wider range in the cultural background of the homes, so Dr Halsey maintains, we should have found a wider range in the children's intelligence; and in that case, particularly had defectives also been included, much larger discrepancies would have been discovered.

However, it can, I think, be safely stated that none of these objections is warranted. Had our critics referred to the original reports they would have seen that many of the children included among the twins reared apart were not in fact pupils at elementary schools. Three pairs were children of parents in the highest professional category; and the two brightest twins were sons of an Oxford don who had died just before they were born (their story is told in full by Conway, 1958, p. 186). Moreover, there was a disproportionately large number from the 'lower professional' category: parents of this class, when the family increased in size, often found it hard to preserve the standard of appearance they thought proper to their station and to give their children the type of education they deserved, since in those days education at a grammar school entailed payment of fees.

In arranging placements it is undoubtedly true that an endeavour is usually made to find foster-parents of the same social class as that of the original parents, but such efforts frequently fail. Parents in the skilled classes are not only reluctant to have their own twins separated, but also disinclined to accept foster children from other families. Parents in the unskilled classes can seldom provide suitable foster homes. On the other hand, childless couples who are well-to-do are often eager to adopt a child; and many of the children thus chosen are healthy and attractive-looking infants from the lower occupational categories. The consequence is that, contrary to Halsey's assumption, the average environmental difference between separated twins is actually greater than it would have been, had the homes been selected purely at random: only nine of the fifty-three separated twins were placed with foster-parents of the same social class as their own parents.

As regards range of intelligence, our group included a couple of mental defectives from special schools (I.Q. 66), and, at the other end of the scale, two scholarship winners (I.Q. 136 and 137)—a difference of 71 points, which is wider than the range one might normally expect in a sample of only fifty-three (namely, 4·5 S.D.). However, for correlations to be comparable, what are really important are not so much the *ranges* as the *standard deviations*. The standard deviation of the group of separated monozygotic twins was 15·3; and that of the ordinary siblings was 15·0. The American groups, it is true, showed a bigger range with standard deviations rising to 17·3; but the correlations now generally printed are nearly always corrected for this disparity (see Woodworth, 1941).

III. Results

Means and correlations

The average intelligence of the twins measured on a conventional I.Q. scale (S.D. = 15) was 97·8 for the separated monozygotes, 98·1 for monozygotes brought up together, 99·3 for the dizygotes as compared with 100·2 for the siblings and 100·0 for the population as a whole. The fact that twins—particularly uniovular—tend to have a lower average ability than ordinary children of the same social background has since been confirmed in several large-scale investigations carried out by Husén (1959), Sandon (1959), and others. The intra-class correlations obtained with the various groups for the different measurements and estimates are set out in Table 2. Since several writers (e.g. Heim, 1954; Hudson, 1965) have stated that the figures reported by Newman *et al.* (1937) for American children imply different conclusions from those drawn in my previous reports, I have appended the corresponding correlations obtained from their inquiry: their raw figures were corrected for age and range by McNemar, and the slight changes this involves have been accepted by Holzinger.

(*a*) *Intelligence*

(1) *Twins reared together.* It will be seen that, both in our own inquiries and in those of Newman, there are large and significant differences between the correlations for 'identical' and for 'non-identical' twins, when both members of the pairs have been brought up in their own homes. Between 'non-identical' twins (at any rate with our own data) the correlations are much the same as those between ordinary brothers and sisters; Newman reports slightly higher figures, as might be expected from the types of test employed. Shields obtained a coefficient of only 0·51 for his dizygotic twins, which is even lower than ours. There is a small positive correlation between the assessments for unrelated children brought up in the same home (0·27). This would seem attributable to the efforts (not very successful) to place these children in foster homes corresponding to those of their parents. Still more recent results, chiefly American, fully confirm our main finding. Using composite scores from tests for 'primary mental abilities' Blewett (1954) obtained correlations of 0·75 and 0·39 for monozygotic and dizygotic twins: but the size of his samples (26 pairs in each group) prevents the difference from being fully significant. Nichols (1965) used composite scores obtained from the National Merit Scholarship Qualifying Tests: pairs differing in school, sex or health records were excluded, and the correlations obtained were 0·87 for monozygotic twins (687 pairs) and 0·63 for dizygotic twins (482 pairs)

(2) *Twins reared apart.* Comparisons of the foregoing type are commonly dismissed on the ground that 'identical' twins naturally keep together much more closely than other siblings. I find it hard to believe that this of itself could account for the large differences observed. However, the study of cases in which twins of identical constitution have been separated almost from birth is likely to carry far more conviction. Unfortunately the number of investigations is extremely small.

In our own set of results the outstanding feature is undoubtedly the high correlation for 'intelligence' between monozygotic twins even when reared apart—0·87 as compared with 0·54 for dizygotic pairs reared together—a difference which is over four times its standard error. With children picked at random the average difference

Table 2. *Correlations for mental, educational, and physical characteristics*

	A. Burt *et al.*						B. Newman *et al.*		
	Mono-zygotic twins reared together	Mono-zygotic twins reared apart	Dizygotic twins reared together	Siblings reared together	Siblings reared apart	Unrelated children reared together	Mono-zygotic twins reared together	Mono-zygotic twins reared apart	Dizygotic twins reared together
Number of pairs*...	95	53	127	264	151	136	50	19	51
Intelligence									
Group test	0·944	0·771	0·552	0·545	0·412	0·281	0·922	0·727	0·621
Individual test	0·918	0·863	0·527	0·498	0·423	0·252	0·881	0·767	0·631
Final assessment	0·925	0·874	0·453	0·531	0·438	0·267	—	—	—
Educational									
Reading and spelling	0·951	0·597	0·919	0·842	0·490	0·545	—	—	—
Arithmetic	0·862	0·705	0·748	0·754	0·563	0·478	—	—	—
General attainments	0·983	0·623	0·831	0·803	0·526	0·537	0·892	0·583	0·696
Physical									
Height	0·962	0·943	0·472	0·501	0·536	−0·069	0·932	0·969	0·645
Weight	0·929	0·884	0·586	0·568	0·427	0·243	0·917	0·886	0·631
Head length	0·961	0·958	0·495	0·481	0·506	0·110	0·910	0·917	0·691
Head breath	0·977	0·960	0·541	0·510	0·492	0·082	0·908	0·880	0·654
Eye colour	1·000	1·000	0·516	0·554	0·524	0·104	—	—	—

* Figures for boys and girls have been calculated separately and then averaged. In columns 3, 4, 5 and 6 the correlations for head-length, head-breadth, and eye colour were based on samples of 100 only.

between their I.Q.s is approximately 17 points, with 'fraternal' twins 12 points, and with 'identical' twins, when reared apart, just under 6 points. The correlations reported by Newman (last 2 columns of Table 2) show similar but somewhat smaller differences. Shields (1962) in his recent publication reports a correlation of 0·77 for separated monozygotes age 8–59 yr (the figure is actually higher than that for monozygotes brought up together 0·76), and a correlation of 0·51 for dizygotes brought up together; but, owing to the small samples, the difference, taken by itself, is too small to be statistically significant. The reader will find it instructive to note how closely the various correlations for intelligence resemble those for most of the physical characteristics. Those for weight, however, are evidently affected by environmental conditions.

In most other researches on twins the investigators have not been in a position to secure the information needed to discriminate between monozygotic and dizygotic pairs. They have therefore merely recorded separate correlations for pairs of like and of unlike sex. Maxwell (1953, p. 144) has collected data from earlier studies; to these we can now add the figures recently reported by Vandenberg (1956), Sandon (1959) and Husén (1959). This yields seven independent researches, based on well over 1000 pairs of like sex and nearly 1000 of unlike. For intelligence the average correlations are 0·76 and 0·57 respectively. Applying the formula given above (p. 141) and using Fisher's z-transformation, the probable averages for the monozygotic and dizygotic pairs respectively work out at 0·89 and 0·56. In a survey just published, Wiseman (1964, p. 99) has computed estimates from data obtained in Manchester schools, using a rather different method; his figures are 0·92 for 'identical' twins and 0·51 for 'fraternal'. All four estimates, it will be observed, are in close agreement with our own, where the zygosity was determined by direct empirical evidence.

(*b*) *Attainments*

The differences obtained with the scholastic tests present a striking contrast with those for intelligence. In our own investigation the correlations for non-identical twins and for siblings reared together (0·83 and 0·80) are far higher than those for identical twins reared apart (0·62) and nearly as high as those for identical twins reared together. (The difference is even larger with Newman's groups.) For siblings reared apart the correlation sinks to 0·53. Here, therefore, the influence of environment is unmistakable. The coefficients which are most conspicuously increased are, as one might expect, those for verbal attainments.

Criticisms

The inferences which my colleagues and I drew from our earlier studies of twins (as well as from other lines of evidence) have been sharply criticized by those who favour an environmentalist theory. But the objections advanced rest almost entirely on armchair arguments: hardly any of the critics cite first-hand evidence of their own. Heim (1954), for example, observes that, even when so-called 'identical' twins have been brought up together, the correlations for intelligence 'still fall short of unity', and she contends that this tells against the genetic hypothesis. But no tests or assessments can claim perfect reliability; and the correlations she cites are almost, if not quite, as high as those obtained from the same individuals on two different occasions. Quoting Newman, she tells us that 'when monozygotic twins are separated,...the differences between their scores are as great as those between unseparated dizygotic twins.' But here there is clearly a mistake. Newman's average difference for dizygotic twins is 9·2 points, but for the separated monozygotic twins only 8·2 points—a reduction which certainly cannot prove the superior importance of environment.

Maddox (1957), Halsey (1959), and Hudson (1965) are still more emphatic in their rejection of genetic influences. Like so many writers, however, they speak of 'differences in *intelligence*' and 'differences in *scores from intelligence tests*' as though the two were synonymous, and tacitly assume that what is true of one type of test—a verbal group test, for instance—holds good of all other assessments. In Newman's investigation the average difference between the separated twins is considerably increased by the big differences obtained in a very small number of pairs. There were four in which the difference amounted to at least 17 points; and in each there was marked difference in the amount of schooling received—ranging from 4 to 15 years. Gladys, for instance, brought up in an 'isolated part of the Rockies where there were no schools', obtained an I.Q. of only 92, whereas her twin sister Helen, who had been to college, obtained an I.Q. of 116—a difference of 24 points. Hudson quotes the case of Mary and Mabel where the difference was 17 points; here there was 4 years difference in schooling. But what does all this prove? Not that the innate intelligence of the twins was widely different, but merely that predominantly verbal tests, like the Otis and the Stanford–Binet, do not furnish very accurate assessments, when the schooling and cultural background are so dissimilar. Newman's own conclusion is very different from that drawn by our various critics. He observes that, even where there was a large discrepancy in the actual scores, it nevertheless seems probable that 'the twin with the lower I.Q. had an *inherited capacity* to reach the rating of the other, had he or she enjoyed the same opportunities'; on the other hand, 'even with a good education the poorly endowed person does not reach the level of a potentially able but poorly educated person' (Newman *et al.* 1937). And in a later summary of his main results he remarks that, 'throughout the whole study of identical twins reared apart', he was, as a geneticist, 'much more impressed with the very great intra-pair *similarities* of the twins, after they had been exposed to all sorts of environmental differences, than with the *differences*' (Gardner & Newman, 1940).

Environmental influences

I have never maintained, as McLeish (1963) implies, that environment has no effect whatever. I have always insisted that genetic and environmental factors are continually interacting from the very start. In our own data the environmental influences are most obvious in the case of group tests. And with every type of test or assessment a *small* portion of the correlation still seems to be attributable to environmental conditions. However, to demonstrate the importance of environmental opportunities it is quite misleading to pick out, as Dr Hudson and other critics have done, just one or two striking instances where a large discrepancy in the test-results is accompanied by a marked difference in home circumstances: once can always counter such an argument with cases, like that of 'George' and 'Llewellyn' (see Conway, 1958, p. 186), where, in spite of an exceptionally wide difference in upbringing, both twins happen to score almost exactly the same I.Q. The only satisfactory method is to correlate the *differences* between the various mental or scholastic assessments for each member of a pair with the *differences* in the material and cultural conditions (see Conway, 1959, pp. 8f.).

For this purpose we have assessed the economic and cultural conditions of the homes in terms of a conventional scale similar to that employed for assessing intelligence and educational attainments, namely, one in which the mean is 100 and the standard deviation is 15. The correlations thus obtained are shown in Table 3.

Table 3. *Correlations of differences for monozygotic twins reared apart*

	Home conditions	
Test results Differences in intelligence	Differences in cultural conditions	Differences in material conditions
Group test	0·43	0·21
Individual test	0·26	0·16
Final assessment	0·15	0·18
School attainments	0·74	0·37

Coefficients over 0·29 are significantly different from zero ($P < 0{\cdot}05$).

It will be seen that differences in educational attainments are highly correlated with differences in the cultural background: there is also a significant correlation between cultural differences and differences in the scores for the group test taken as they stand. But the correlations for the individual test and for the final assessment are so low as to be non-significant with a sample of this size. The differences in educational attainments show a small but significant correlation with differences in material conditions, chiefly no doubt because the latter are responsible for differences in the children's physical health and school attendance.

The multifactorial hypothesis

Most of the critics who seek to belittle the importance of mental inheritance seem content to examine just one particular investigation and base their arguments on the weaknesses they discern in that. They ignore the fact that the hereditarian's conclusions rest, not on a single research however elaborate, but on inferences drawn from a wide variety of different approaches—all leading to a systematic theory of mental inheritance, very similar to that arrived at by contemporary geneticists working in entirely different non-mental fields.

Accordingly, it seems desirable by way of conclusion to compare the correlations obtained from twins and siblings with those obtained from other pairs related by various degrees of affinity. The figures compiled from our London surveys are set out in Table 4. Since our last publication, further data have been collected for some of the smaller groups, and in these cases the correlations have been recalculated: the figures show little change, but their standard errors are smaller. We have endeavoured to select the individuals composing the groups so that they should, as far as possible, be genuinely representative of the population as a whole. With the smaller groups this has not been easy; but for all of them the standard deviations, which range from just under 14 to a little over 16 I.Q. points, are much the same. To compare or correlate figures for adults with those obtained from children may seem a questionable procedure. But many of the pupils tested between 1914 and 1924 have since grown up; and it has often been possible to trace and re-test them as well as to test their own children. In that way correlations between assessments for these children and similar assessments secured from their parents when they were children can be computed. On the whole, they fully confirm the values obtained in the usual way.

Table 4. *Correlations between relatives*

	Burt		Other investigators		
	Number of pairs	Correlation	Number of investigations	Median correlation	Theoretical value
Direct line					
With parents (as adults)	374	0·49	13	0·50	0·49
With parents (as children)	106	0·56	—	—	0·49
With grandparents	132	0·33	2	0·24	0·31
Collaterals					
Between monozygotic twins					
Reared together	95	0·92	13	0·87	1·00
Reared apart	53	0·87	3	0·75	1·00
Between dizygotic twins					
Same sex	71	0·55	8	0·56	0·54
Different sex	56	0·52	6	0·49	0·50
Between siblings					
Reared together	264	0·53	36	0·55	0·52
Reared apart	151	0·44	33	0·47	0·52
Between uncle (or aunt) and nephew (or niece)	161	0·34	—	—	0·31
Between first cousins	215	0·28	2	0·26	0·18
Between second cousins	127	0·16	—	—	0·14
Unrelated persons					
Foster parent and child	88	0·19	3	0·20	0·00
Children reared together	136	0·27	4	0·23	0·00
Children reared apart	200	−0·04	2	−0·01	0·00

For purposes of comparison I also give medians for results reported by other investigators and the hypothetical values to be expected in accordance with what may be called the neo-Mendelian theory of mental inheritance. To calculate the latter the method used is a modification of that originally develop by Fisher (1918) in his study of body measurements. The formulae (Burt & Howard, 1956, pp. 115–16), unlike those used by most other writers, allow for two facts commonly ignored: in the case of intelligence there is (i) ample evidence for assortative mating between parents, and (ii) some evidence for a slight but incomplete amount of dominance. To some extent these two opposing influences tend to neutralize each other; but the net result is usually to raise the theoretical figures somewhat above those generally published, which tacitly assume that dominance is absent and that mating is perfectly random. In that case with simple multifactorial inheritance the theoretical correlations would be—with parents 0·5, with grandparents 0·25, with cousins 0·125, i.e. $(\frac{1}{2})^n$ for correlations between relatives of the *n*th degree. With intelligence we believe that unifactorial as well as multifactorial modes of transmission operate. And slight indications of sex-linkage, chiefly associated with the former, also seem discernible, particularly in the correlations for dizygotic twins. However, both these further influences, if they operate at all, are relatively small, and were consequently neglected in deducing the formulae used.

Erlenmeyer-Kimling & Jarvik (1964) have reported the results of a somewhat similar review of the literature. As they point out, 'while behaviour theory as yet makes few provisions for modern genetic concepts, the literature of the past half-century contains far more information than is generally realized about the relation between genotypic similarity and performance on mental tests'. They have located fifty-two such studies yielding 'over 30,000 correlational pairings from 8 countries in 4 continents', and present the main results in the form of an instructive diagram, which exhibits according to the degree of relation every correlation in their collection, together with the median for the various groups. Within some of the groups, particularly those where members of the same family have been brought up together, the coefficients exhibit a fairly wide range, e.g. for siblings they range from just over 0·30 to a little over 0·70. The reasons are fairly obvious. The low values have usually been obtained with somewhat unreliable tests, occasionally from but a single unvalidated test; often, too, the groups are fairly homogeneous. The high values are generally derived from heterogeneous samples, varying considerably in home background and education, and the tests are frequently verbal group tests, such as would be appreciably affected by differences in education. For most of the comparisons, however, there is a clear concentration of values near the mean or median. My own list includes a number of studies, chiefly British, which do not appear in the American collection; and with the writers' permission I have now added to my own earlier compilation data from several inquiries which they quote and which had escaped my notice.

In Table 4 I have given median values only; and I have indicated the number of investigations on which the medians are based. It will be seen that the values reported by other investigators tend to be somewhat higher than my own, presumably because many based their correlations on scores obtained from tests of a predominantly verbal type. If allowance is made for this, then, it will be seen that (except perhaps in the case of first cousins in my own research) both the figures obtained in the London inquiry and the medians of the figures published by other observers are in close agreement with the theoretical values we have calculated according to the neo-Mendelian hypothesis.

IV. SUMMARY AND CONCLUSIONS

1. Various mental and physical characteristics have been measured for persons related in different degrees, and the correlations compared both with each other and with those obtained by other investigators, as well as with the values to be expected in accordance with the modified theory of multifactorial inheritance outlined in an earlier paper (Burt & Howard, 1956). In particular, an intensive correlational study has been made of monozygotic twins reared together and apart.

2. For assessments of 'intelligence' the correlations from unrelated pairs of children brought up together are positive but comparatively small; those from related pairs increase progressively according to the closeness of family relation. The correlation for monozygotic twins reared in separate environments amounts to as much as 0·88, and is but slightly increased where such twins have been reared together. On the other hand, the correlations for school attainments vary closely with similarity in environmental circumstances. From this it may be inferred that individual

differences in 'intelligence', particularly when the assessments have been carefully checked, are influenced far more by genetic constitution, or what is popularly termed 'heredity', than by post-natal or environmental conditions.

3. The detailed values for the different groups obtained both in my own investigations and in those of other writers, agree satisfactorily with the view that the genetic tendencies which are responsible for individual differences in intelligence are in the main (though probably not exclusively) transmitted in accordance with a multifactorial or polygenic hypothesis, due allowance being made for assortative mating and partial dominance.

References

Blewett, D. B. (1954). An experimental study of the inheritance of intelligence. *J. ment. Sci.* **100**, 922–33.

Burt, C. (1912). The inheritance of mental characteristics. *Eugen. Rev.* **4**, 168–204.

Burt, C. (1921). *Mental and Scholastic Tests*, 4th ed. London: P. S. King, Staples Press (1962).

Burt, C. (1933). *Handbook of Tests for Use in Schools*. London: Staples Press.

Burt, C. (1943). Ability and income. *Brit. J. educ. Psychol.* **13**, 83–98.

Burt, C. (1955). The evidence for the concept of intelligence. *Brit. J. educ. Psychol.* **25**, 158–77.

Burt, C. (1958). The inheritance of mental ability. *Amer. Psychologist*, **13**, 1–15.

Burt, C. & Howard, M. (1956). The multifactorial theory of inheritance and its application to intelligence. *Brit. J. statist. Psychol.* **9**, 95–131.

Burt, C. & Howard, M. (1957). Heredity and intelligence a reply to criticism. *Brit. J. statist. Psychol.* **10**, 33–63.

Carr-Saunders, A. M. & Jones, D. C. (1937). *A Survey of the Social Structure of England and Wales*. Oxford: Clarendon Press.

Conway, J. (1958). The inheritance of intelligence and its social implications. *Brit. J. statist. Psychol.* **11**, 171–90.

Conway, J. (1959). Class differences in general intelligence. *Brit. J. statist. Psychol.* **12**, 5–14.

Darlington, C. D. (1963). Psychology, genetics and the process of history. *Brit. J. Psychol.* **54**, 292–9.

Darlington, C. D. & Mather, K. (1949). *The Elements of Genetics*. London: Allen and Unwin.

Erlenmeyer-Kimling, L. & Jarvik, L. F. (1964). Genetics and intelligence. *Science* **142**, 1477–9.

Essen-Möller, E. (1941). Empirische Ähnlichkeitsdiagnose bei Zwillingen. *Hereditas* **17**, 1–23.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edin.* **52**, 399–433.

Galton, F. (1883). *Inquiries into Human Faculty*. London: Macmillan.

Gardner, I. C. & Newman, H. H. (1940). Mental and physical traits of identical twins reared apart. *J. Hered.* **31**, 119–26.

Gaw, F. (1925). *Performance Tests of Intelligence*. M.R.C. Report no. xxi. London: H.M.S.O.

Halsey, A. H. (1959). Class differences in intelligence. *Brit. J. statist. Psychol.* **12**, 1–4.

Heim, A. W. (1954). *The Appraisal of Intelligence*. London: Methuen.

Hudson, L. (1965). A curious dispute: inheritance versus environment. *The Listener* **73**, 929–32.

Husén, T. (1959). *Psychological Twin Research*. Stockholm: Almquist and Wiksele.

Juel-Nielsen, N. & Mogensen, A. (1957). Uniovular twins brought up apart. *Acta Genet.* **7**, 430–3.

Lewis, D. G. (1957). The normal distribution of intelligence: a critique. *Brit. J. Psychol.* **48**, 98–104.

Maddox, H. (1957). Nature–nurture balance sheets. *Brit. J. educ. Psychol.* **27**, 166–75.

Maxwell, J. (1953). *Social Implications of the* 1947 *Scottish Mental Survey*. London: University of London Press.

McLeish, J. (1963). *The Science of Behaviour*. London: Barrie and Rockliff.

Muller, H. J. (1925). Mental traits and heredity. *J. Hered.* **16**, 433–48.

Newman, H. H., Freeman, F. N. & Holzinger, K. J. (1937). *Twins: A Study of Heredity and Environment*. Chicago University Press.

NICHOLS, R. C. (1965). The inheritance of general and special ability. *National Merit Scholarship Corporation Research Report* **1**, 1.

PENROSE, L. S. & SMITH, S. M. (1955). Monozygotic and dizygotic twin diagnosis. *Ann. hum. Genet.* **19**, 273–89.

POPENOE, P. (1922). Twins reared apart. *J. Hered.* **5**, 142–4.

SANDON, F. (1959). Twins in the school population. *Brit. J. statist. Psychol.* **12**, 133–8.

SHIELDS, J. (1962). *Monozygotic Twins*. London: Oxford University Press.

SPIEGELMAN, S. (1946). Nuclear and cytoplasmic factors controlling enzymatic constitution. *Lond. Sch. Hyg. Symp. quant. Biol.* **11**, 256–77.

STOTT, D. H. (1956). *Unsettled Children and their Families*. London: University of London Press.

VANDERBERG, S. G. (1962). Innate abilities. *Louisville Twin Study Research Report*, no. 3. University of Louisville, Ky.

WISEMAN, S. (1964). *Education and Environment*. Manchester University Press.

WOODWORTH, R. S. (1941). Heredity and environment: a critical survey of recently published material on twins and foster children. *Soc. Sci. Res. Council Bull.* no.47.

WOOLF, B. (1952). Appendix to *Quantitative Inheritance*, ed. E. C. R. Reeve and C. H. Waddington. London: H.M.S.O.

(*Manuscript received* 8 *September* 1965)

Brit. J. Psychol (1966), 57, 1 *and* 2, *pp.* 155–159
Printed in Great Britain

PERSONALITY AND JUDGEMENTS OF TEMPORAL INTERVALS

BY C. R. BELL AND ANNE N. WATTS
Medical Research Council Environmental Physiology Research Unit,
London School of Hygiene and Tropical Medicine

The ways in which subjects performed a number of time-estimation tasks were examined for associations with normal pulse rate, normal oral temperature, age, sex, intelligence and personality characteristics. No universal relations were found in the data, though there was evidence to suggest that age, pulse rate, oral temperature and some aspects of personality possibly are worthy of further research to determine whether they might be related to some types of time estimation.

Previous research into relations between time estimation and individual characteristics has been inconclusive. Thor (1962) found consistent diurnal effects in production estimates of 30 sec and 120 sec intervals which 'point to the influence of an underlying physiological variable'. Hoagland (1933, 1935, 1951) and Baddeley (1965) have shown relations between estimates of a counting rate of 1 digit/sec and induced changes in body temperature. Other workers (Bell, 1965; Bell & Provins, 1963; Shaeffer & Gilliland, 1938) have failed to show consistent relations between time estimation and changes in physiological parameters. Gilliland & Humphreys (1943) and Smythe & Goldstone (1957) have reported an age effect on the variability of time judgements when the performance of children and adults was examined. MacDougall (1904) and Yerkes & Urban (1906) found differences between the sexes in their estimations in the range 18–108 sec, but Orme (1962) concluded that there is no consistent evidence for a sex effect.

Of the studies of the relations between time judgements and personality measures in non-clinical samples, Loehlin (1959) reported that no relation emerged between MMPI scores and the verbal estimates of intervals from 1 sec to 20 min of seventy-four subjects. Other studies (Hare, 1963; Siegman, 1962) have shown conflicting results.

In the present report, data from a non-clinical group of subjects who were asked to make temporal judgements of various magnitudes by various methods of estimation are compared with the normal (resting) oral temperature and pulse rates, age, sex and personality characteristics of the subjects.

METHODS

The data for analysis came from six investigations, I–VI. In no investigation did the subjects have knowledge of results.

The investigations

I. Eight volunteer naval ratings made production estimates of 4, 9 and 13 min. All subjects had completed the Heron Two-Part Personality Measure (Heron, 1956) as part of the investigation.

II. Eighteen volunteer male subjects, scientific, military and technical personnel, completed the Heron Personality Measure. During performance on a five-choice serial reaction time task, fourteen made a verbal estimate of 4 min, eighteen made a verbal estimate of 9 min, and sixteen made a verbal estimate of 13 min.

III. Ten volunteer male shop assistants completed the Heron Personality Measure and, whilst sitting at rest, each estimated intervals of 45 and 100 sec by counting to himself at a rate he estimated to be 1 digit/sec. The subject operated a covered stopwatch during his count which recorded the time taken.

IV. Thirty volunteer naval ratings, over a period of 5 days, completed the Heron Personality Measure, the 16 PF scale (Cattell, Saunders & Stice, 1957), the *F* scale of the Authoritarian Personality Study (Adorno, Frenkel-Brunswik, Levinson & Sanford, 1950) and the IPAT self-analysis form (Cattell, 1957). Data were also available of the subjects' age, resting pulse rate, resting oral temperature and scores on the Wechsler Adult Intelligence Scale. Each subject estimated an interval of 60 sec by counting to himself at a rate of 1 digit/sec. The time taken was recorded on a covered stopwatch which the subject operated.

V. Thirty-five mature students (male and female) attending evening classes in Psychology completed the 16 PF scale and gave their date of birth. The following week they estimated a rate of counting at 1 digit/sec over a period of 60 sec by recording the number of digits they had counted between the experimenter saying 'Go' and 'Stop'.

VI. Forty-seven volunteer male college students individually completed the Heron Personality Measure, the 16 PF scale, and the IPAT self-analysis form. Their normal (resting) oral temperature and normal (resting) pulse rate were recorded. The date of birth of each subject was noted. Estimations of a counting rate of 1 digit/sec for intervals of 60 sec were made by the methods used in IV and V above. The subjects tapped a morse key at an estimated rate of 3 taps/sec, and during completion of the 16 PF questionnaire made verbal estimates of 10 min and 20 min. Forty-five of the subjects judged whether or not a metronome device was flashing at a rate of 1 flash/sec during the random presentation of rates of flashing within the range of 20–180 flashes/min. The order of presentation of the time estimation tasks and the personality measures was randomized over the entire experiment except that two time estimation tasks or two personality questionnaires were never given successively.

Treatment of data

The time-estimation data from all investigations were converted to standard scores so that a score of 100 indicated a perfect estimation. On each time-estimation task the subjects were grouped into the 18 % who had the largest converted time-estimation scores on the task (fast counters), the 18 % who had the smallest converted time-estimation scores on the task (slow counters) and the intermediate 64 %. Where estimates were tied, the 18 % extreme was increased to incorporate subjects who produced the same estimate.

In all examinations of the association between the time-estimation data and the oral temperature, pulse rate, age intelligence, and personality variables where data were gathered from more than ten subjects, *t* tests were used to compare the characteristics of the fast, intermediate and slow counters.

Linear regression coefficients and product–moment correlations were calculated in respect of all the converted and log-transformed time-estimation data on all the personality data available. Linear regression coefficients were also calculated for the regressions of the raw time-estimation data on personality data in respect of the production task estimates of 4, 9 and 13 min from investigation I and of the stopwatch task estimates of 60 sec from investigations III, IV and VI. Quadratic regression coefficients were similarly calculated. Analyses of variance were used to test all the coefficients for significant departure from zero.

RESULTS

The comparisons of the individual characteristics of the fast, intermediate and slow counters involved the calculation of 440 *t*-tests of which only 20 (4·5 %) attained statistical significance at $P < 0{\cdot}05$.

Pulse-rate measurements were lower for fast than intermediate counters on the estimated 1 digit/sec counting task and 3 taps/sec task, and were lower for slow than for intermediate counters on the 1 digit/sec counting task. The normal oral temperature of slow counters in the verbal estimation of 10 or 20 min was lower than that of the intermediate counters. On part I of the Heron Personality Measure, fast counters in verbally estimating 10 or 20 min were emotionally less adjusted than intermediate counters. Fast counters on the 3 taps/sec task were emotionally less stable (16 PF, *C* scale) and slow counters on the 1 digit/sec stopwatch task were emotionally more stable (16 PF, *C* scale) than intermediate counters on the same tasks. Fast counters on the 1 digit/sec stopwatch task were less dominant (16 PF, *E* scale) than either intermediate or slow counters. Intermediate counters were less guilt prone (16 PF, *O* scale) than either fast or slow counters when verbally estimating 10 or 20 min and slow counters were less guilt prone than fast counters when estimating 1 digit/sec on the stopwatch task. Slow counters were less tense and excitable (16 PF, Q_4 scale) than either fast or intermediate counters on the 1 digit/sec stopwatch task and than fast counters on the 3 taps/sec task. Intermediate counters were less sensitive and effeminate (16 PF, *I* scale) than fast counters and more sophisticated and polished (16 PF, *N* scale) than slow counters on the perception of 1 flash/sec task, were more enthusiastic and happy-go-lucky (16 PF, *F* scale) than fast counters on the verbal estimation of 10 or 20 min task and were less bohemian, introverted, absent-minded (16 PF, *M* scale) than slow counters in both the perception of 1 flash/sec and verbal estimation of 10 and 20 min tasks.

Of the 302 linear and quadratic regression analyses of the raw, converted and log-transformed time-estimation data on the individual characteristics of the subjects, in only nineteen (6·3 %) did the regression account for a significant proportion of the variance in the data and depart significantly from zero. These regressions indicated an association between both rate and log rate of tapping at an estimated 3 taps/sec and pulse rate, of both a linear ($r = -0{\cdot}388$ and $-0{\cdot}383$ respectively) and a quaratic form. A log transformation of the verbal estimates of 10 and 20 min was linearly related ($r = +0{\cdot}292$) to normal oral temperature. The log transformed rate of counting at an estimated 1 digit/sec was related quadratically to the age of the subjects in a sample with an age-range of 16–58 yr. Speed of tapping at an estimated 3 taps/sec was shown to have a quadratic regression on the covert anxiety score of the IPAT self-analysis form. Of the personality characteristics as assessed on the 16 PF questionnaire, six were significantly related to the time-estimation data. On cyclothymia (16 PF, *A* scale) there was a significant linear regression ($r = -0{\cdot}231$) of log counting at an estimated rate of 1 digit/sec, and significant quadratic regressions of the converted and log-transformed data from the 1 digit/sec stopwatch task. General intelligence (16 PF, *B* scale) was related to verbal estimates of 10 and 20 min ($r = -0{\cdot}309$). Both the converted and log-transformed data from the 1 digit/sec stopwatch task produced significant linear regressions ($r = -0{\cdot}262$ and $-0{\cdot}253$) on the dominant–submissive scale (16 PF, *E* scale). Speeds of tapping at an estimated 3 taps/sec for both converted and log-transformed data were shown to have quadratic regressions on the 16 PF, *I* scale (Premsia). A positive linear regression ($r = +0{\cdot}219$) was found of log speed of counting at an estimated 1 digit/sec on the shrewdness, sophistication scale (16 PF, *N* scale). Finally, scores of ergic tension (16 PF, Q_4 scale) indicated significant positive linear regressions of tapping rate at an estimated 3 taps/sec ($r = +0{\cdot}291$), and of converted and log-transformed estimates of 1 digit/sec counting rate on the stopwatch task ($r = +0{\cdot}255$ and $+0{\cdot}277$) on the ergic-tension-scale scores.

The estimates of a rate of counting at 1 digit/sec were compared between the thirteen male evening-class students and twenty-two female evening-class students in investigation V. There was no difference between the two groups in their mean estimates. In the regression analyses, only the significant negative linear regression of the counting at 1 digit/sec estimation task on parmia (16 PF, H scale) was significantly different from the relation between the two parameters in the male subjects.

Discussion

The primary finding of the present investigations was perhaps that the number of significant associations between time estimation and personality, age, sex and physiological parameters did not differ significantly from chance. Those significant results which did emerge were not confined to one or two time-estimation tasks or to one or two of the twenty-five parameters examined in the analyses, but were distributed throughout the tasks and the variables. Thus, the question of a relation between time judgements and physiological measurements, age and sex appears to be unresolved.

Though personality characteristics, or psychiatric assessments, of clinical groups have been found to be associated with time judgements of several kinds (Orme, 1962), the basis of a relation between non-clinical personality characteristics and time judgements has eluded research workers. The perception, experience and demonstration of time-estimation skills is probably as complicated a topic as many other areas of psychology (Gilliland, Hofeld & Eckstrand, 1946; Wallace & Rabin, 1960). Schilder (1936) suggested, 'Perception and experiences of any other type are in no way stabilized entities. The organism...is a psychical organism with emotions and drives which determine the final structure of our perceptive world. Time perception will be dependent upon the same factors.' At present, the identification of these factors in studies of time judgements has not been completed.

Conclusion

When subjects had to make temporal judgements on a number of time estimation tasks, the individuality of their responses was not consistently related to any of the variables studied. There appears to be some suggestion in the present data that the variables of resting pulse rate and oral temperature, age, sex and certain personality characteristics may be worthy of further exploration.

The authors wish to acknowledge their gratitude to the Royal Naval Personnel Research Committee for facilitating parts of these investigations.

References

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J. & Sanford, R. N. (1950). *The Authoritarian Personality*. New York: Harper.

Baddeley, A. D. (1965). Reduced body temperature and time estimation. *Bull. Brit. psychol. Soc.* **18**, no. 58, 2A.

BELL, C. R. (1965). Time estimation and increases in body temperature. *J. exp. Psychol.* **70**, 232–4.

BELL, C. R. & PROVINS, K. A. (1963). The relation between physiological responses to environmental heat and time judgments. *J. exp. Psychol.* **66**, 572–9.

CATTELL, R. B. (1957). *Handbook for the IPAT Anxiety Scale.* Champaign, Ill.: Inst. Personality and Ability Testing.

CATTELL, R. B., SAUNDERS, D. R. & STICE, G. (1957). *Handbook for the Sixteen Personality Factor Questionnaire.* Champaign, Ill.: Inst. Personality and Ability Testing.

GILLILAND, A. R., HOFELD, J. & ECKSTRAND, G. (1946). Studies in time perception. *Psychol. Bull.* **43**, 162–76.

GILLILAND, A. R. & HUMPHREYS, D. W. (1943). Age, sex, method and interval as variables in time estimation. *J. genet. Psychol.* **63**, 120–30.

HARE, R. D. (1963). Anxiety, temporal estimation and rate of counting. *Percept. mot. Skills* **16**, 441–4.

HERON, A. (1956). A two-part personality measure for use as a research criterion. *Brit. J. Psychol.* **47**, 243–51.

HOAGLAND, H. (1933). The physiological control of judgments of duration: evidence for a chemical clock. *J. gen. Psychol.* **9**, 267–87.

HOAGLAND, H. (1935). *Pacemakers in Relation to Aspects of Behaviour.* New York: MacMillan.

HOAGLAND, H. (1951). In *Problems of consciousness*, ed. H. Abramson. New York: Macy.

LOEHLIN, J. C. (1959). The influence of different activities on the apparent length of time. *Psychol. Monogr.* **73**, no. 474 (whole number).

MACDOUGALL, R. (1904). Sex differences in the sense of time. *Science* N.S. **19**, no. 487, 707–8.

ORME, J. E. (1962). Time studies in normal and abnormal personalities. *Acta psychol., Amsterdam* **20**, 285–303.

SHAEFFER, V. G. & GILLILAND, A. R. (1938). The relation of time estimation to certain physiological changes. *J. exp. Psychol.* **23**, 545–52.

SCHILDER, P. (1936). Psychopathology of time. *J. nerv. ment. Dis.* **83**, 530–45.

SIEGMAN, A. W. (1962). Anxiety, impulse control, intelligence and the estimation of time. *J. clin. Psychol.* **18**, 103–5.

SMYTHE, E. J. & GOLDSTONE, S. (1957). The time sense: a normative genetic study of the development of time perception. *Percept. mot. Skills* **7**, 49–59.

THOR, D. H. (1962). Diurnal variability in time estimation. *Percept. mot. Skills* **15**, 451–4.

WALLACE, M. & RABIN, A. I. (1960). Temporal experience. *Psychol. Bull.* **57**, 213–36.

YERKES, R. M. & URBAN, F. M. (1906). Time-estimation in its relations to sex, age, and physiological rhythms. *Harvard Psychological Studies* **2**, 405–30.

(*Manuscript received* 30 *June* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 161–170
Printed in Great Britain

SOCIAL DESIRABILITY IN INDIVIDUAL PERFORMANCE ON THIRTEEN MMPI SCALES

BY ERLING E. BOE*
University of Washington

AND WILLIAM S. KOGAN†
Veterans Administration Hospital, Seattle

The degree to which individual subjects tend to give socially desirable responses in self-description on each of thirteen MMPI scales was determined by computing, for each subject, a point–biserial correlation between his true/false responses and the social desirability (SD) scale values of the items. The results indicated that, on the average, from 10 % to 20 % of the variance in individual self-descriptions was associated with SD scale values. In comparison with these results, individual differences in the tendency to give SD responses, as measured by (*a*) these point–biserial correlations and (*b*) Edwards's SD scale, account for much more of the variance in raw scores on many of the MMPI scales. The differences between these methods of appraising SD aspects of test performance and implications for constructing SD-free scales were considered.

The social desirability (SD) variable has two distinct meanings (Edwards, 1957, p. vi): it refers (*a*) to the SD scale values (SDSV) of personality statements, and (*b*) to the response tendency of individuals (SD response tendency) to endorse personality statements with socially desirable scale values and to reject statements with socially undesirable (SUD) scale values. In the first sense, the SD variable is an attribute of statements commonly found in personality inventories; in the second sense, it refers to a personality trait.

The SD response tendency itself has two quite distinct meanings. Edwards (1957) observed some time ago that the endorsement of a socially desirable statement or the rejection of a socially undesirable statement could be either genuinely self-descriptive or it could represent SD faking. Block (1965) understands the SD response tendency to mean SD faking exclusively, and has presented a detailed critique of this meaning of the concept. By contrast, Edwards, in a recent communication, regards the SD response tendency as a general personality trait, and does not identify 'this trait with another trait which might be defined as the tendency to falsify responses in self-description'. Although Edwards, Diers & Walker (1962) have presented some evidence indicating that faking-good accounts for only a small part of the variance in MMPI performance, it is not yet possible to separate the two components of SD empirically because no operations have been developed to measure each one separately in a convincing way. Our position is that the SD response tendency, as it is typically measured at the present time, represents a combination of a genuinely self-descriptive component and a faking component. For the purposes of the present research, however, the issue of faking is not germain to the problems investigated. Consequently, traditional techniques of measuring the SD response tendency are employed

---

[*] This work was carried out while the senior author was at the Veterans Administration Hospital, Seattle, Washington. He is now at Brown University, Providence, R.I. 02912.

[†] Also affiliated with the Department of Psychiatry, University of Washington School of Medicine.

in an effort to clarify various issues in assessing the degree to which individual subjects respond in a socially desirable manner on specific scales of a widely used personality inventory.

One issue in SD research that has generated a great deal of controversy is the degree to which SD considerations are involved in the performance of subjects taking personality inventories under standard conditions. Edwards (1961), for example, has contended that 'scores on many personality scales of the True–False type could, in large part, be accounted for by individual differences in the tendency of subjects to give socially desirable responses to items in the scales' (p. 351), while Taylor (1959) concluded that his research 'argues against a simplistic interpretation of the MMPI as being, in large part, a measure of social desirability' (p. 516). The source of these conflicting conclusions can be attributed to methodological differences. The purposes of the present research are first to clarify the methodological differences, and then to subject both methods to a systematic empirical analysis. Since most SD research has been with the MMPI, attention will be focused on it specifically.

One method of determining the extent of SD involvement in inventory performance is based on SDSV, the first meaning of the SD variable. It was first shown (Edwards, 1953; Hanley, 1956; Rosen, 1956; Taylor, 1959; Wiggins & Rumrill, 1959) that SDSVs correlated about 0·84 with probabilities of item endorsement computed from the performance of a group of subjects. While true for grouped data, several researchers (Block, 1965; Boe & Kogan, 1963; Taylor, 1959) later pointed out that these correlations do not accurately represent the SD response tendency in the individual subject. In order to ascertain the extent to which the individual subject's responses given in self-description were associated with SDSVs, these researchers correlated the true–false responses of each subject to a sample of items with the SDSVs of the items using the point–biserial correlation (r_{pb}). They found that SDSVs account, on the average, for 10–30 % of individual MMPI performance. It was from data such as these that Taylor (1959) concluded that the MMPI is not, in large part, a measure of SD. A defining characteristic of this method is that it measures the extent to which individual differences in responding *within* subjects are associated with SDSVs. That is, there is variability (from true to false) in each subject's responses to a sample of items, and it is this variability within individuals that is being correlated with the SDSV meaning of SD. A second characteristic of this method is that it uses dichotomous responses to individual items instead of a single score for a personality scale (i.e. for a group of items).

The size of r_{pb} varies considerably from subject to subject. Since each r_{pb} is literally a measure of the degree to which an individual endorses SD items and rejects SUD items, this method is one way of measuring the SD response tendency, the second meaning of the SD variable (Boe & Kogan, 1964). Differences in the size of r_{pb} indicate individual differences *between* subjects in the SD response tendency.

A second method of determining the extent of SD involvement in inventory performance is based on the second meaning of the SD variable, the SD response tendency. An SD scale, such as Edwards's (1957) *SD* scale, Wiggins's (1959) *Sd* scale, or Crowne and Marlow's (1960) *MC–SD* scale, is first derived, by various procedures, in which SD items are keyed 'true' and SUD items are keyed 'false'. The number of keyed responses is taken as a measure of a subject's SD response tendency and then correlated with scores obtained on personality scales. Scores on Edwards's *SD* scale, for example, have been found to correlate very highly with raw scores on many MMPI scales (Edwards, 1957) and have been repeatedly demonstrated to be a nearly pure measure of the first MMPI factor (Block, 1965; Edwards & Diers, 1962; Edwards *et al.* 1962; Edwards & Heathers, 1962; Edwards & Walker, 1961; Edwards & Walsh, 1963). It was from data such as these that Edwards (1961) concluded that scores on many personality scales could be largely accounted for by SD considerations. In contrast with the r_{pb} technique (which appraises individual differences *within* subjects), this method is characterized by the fact that it appraises the extent to which individual differences *between* subjects in the SD response tendency are associated with scores on personality scales. Thus, it is not surprising that apparently contradictory results can be found when the variables being correlated are different.

Since these two methods of evaluating SD considerations in performance on personality inventories are so different, each should contribute something unique to a better understanding of the SD aspects of responses given in self-description. However, no previous research has been designed to utilize and compare these two methods for the same set of items, and the responses of one sample of subjects. Furthermore, the advantages of applying the r_{pb} technique to samples of items included in specific MMPI scales have not yet been exploited. The present research was designed to investigate these (and other) problems in assessing the SD aspects of individual performance on thirteen MMPI scales commonly used in clinical practice.

METHOD

Subjects and procedure

Two groups of male psychiatric patients who had been admitted to any of the three psychiatric wards of the Veterans Administration Hospital, Seattle, were the subjects. The various measures of SD were computed from the performance of subjects in group I ($n = 50$); the reliability of a single subject in responding to items of the MMPI scales was obtained separately for each subject in group II ($n = 12$).

The subjects in group I were selected by means of a table of random numbers from the population of patients (about 350) who had satisfactorily completed routine intake testing (MMPI, Draw-a-Person Test, and Shipley) upon admission to the psychiatric wards during a recent 1-year period. All newly admitted psychiatric patients are given these tests in small groups, except for those who are too disturbed or resistive to complete them satisfactorily. The median number of days after admission for testing was 4·4; the median age of the sample was 40·5 years.

The subjects in group II were selected on the basis of availability for retesting either 3 or 4 days after routine intake testing, not randomly as were the subjects in group I. Only the MMPI was given during the retest. The tests were administered to small groups of subjects on both occasions.

Statistical analyses

(i) *Correlations with SDSVs.* The association of individual differences within subjects in self-descriptive responses (i.e. true or false responses) with SDSVs (as reported by Dahlstrom & Welsh, 1960) was determined by computing r_{pb} for each subject on each of the three MMPI validity scales, the eight standard clinical scales, the *Mf* scale, and Edwards's *SD* scale. Thus, thirteen r_{pb} were computed for each subject. This method of measuring the SD response tendency has the advantages of (*a*) providing a direct measure of the degree to which individual performance on each MMPI scale is associated with the SDSVs, and (*b*) indicating, by its sign, whether the responses are of a SD or a SUD nature.

(ii) *Correlations with SD response tendency.* For the analysis of individual differences between subjects, raw scores on each of the thirteen MMPI validity and clinical scales were correlated (product–moment correlation) with (*a*) scores on Edwards's *SD* scale and with (*b*) the individual r_{pb} described in the preceding paragraph.

(iii) *Reliabilities.* Because the comparisons of correlations undertaken in this study would be affected if the measures being correlated were of substantially different reliability, reliabilities were obtained so that correlations could be corrected for attenuation. To correct the r_{pb} measure for attenuation, the reliability of the SDSVs and the reliability of individual responses to items of each MMPI scale were needed. Boe & Kogan (1963) and Edwards (1963) have reported test–retest reliability coefficients of 0·97 for SDSVs. Test–retest reliability coefficients of individual responses to items of each MMPI scale were obtained from the performance of group II. For each subject, true/false responses to items of an MMPI scale were arranged in a 2 × 2 contingency table with one dimension representing true/false responses on the first administration and the other dimension representing true/false responses on the second administration. The phi correlation (r_ϕ) computed from this tabulation represented the reliability of individual MMPI responses. (Since the coefficients, r_ϕ, were computed for each subject individually, the n for each coefficient is not the number of subjects in group II, but the number of items in each MMPI scale, which range from 15 (L) to 78 (Sc).) To correct the correlations of the *SD* scale with other MMPI scales for attenuation, the reliability of each scale was needed. Edwards (1957) has reported a corrected split-half reliability coefficient of 0·83 for the *SD* scale, while Rosen (1953) has reported test–retest reliability coefficients for the other MMPI scales which were computed from a sample of male psychiatric patients in a Veterans Administration Hospital.

RESULTS

Correlations with SDSVs

For group I, the algebraic mean r_{pb} between responses given in self-description by individual subjects to the items of each of the thirteen MMPI scales and the SDSVs of those items are reported in row 1 of Table 1. (For all analyses in which correlations were used as scores and then subjected to further statistical analyses, the correlations were transformed to Fisher's z' first.) The same mean r_{pb} for group II is presented in row 2 of Table 1. It is apparent that mean correlations for group II are nearly the same as those for group I, even though the two groups were selected by different criteria.

The standard deviations of the fifty r_{pb} (transformed to z'), whose means are reported in row 1 of Table 1, for each MMPI scale are presented in column 2 of Table 2. It is apparent that individual differences in the SD response tendency were quite large, and that many r_{pb} were negative. A negative r_{pb} indicates the degree of a SUD response tendency. The percentage of group I subjects exhibiting a statistically significant SD response tendency and a statistically significant SUD response tendency on each MMPI scale is tabulated in column 4 and 5, respectively, of Table 2.

In some scales, such as *K* and *Hs*, the SUD tendency is as great as the SD tendency. In these cases particularly, an algebraic mean of the r_{pb} would approach zero (because negative r_{bp} would balance positive r_{pb}) and imply that individual performance was associated only to a slight extent with the SDSVs, on the average. However, such an inference is misleading, and can be corrected by computing a modular mean r_{pb} (disregarding the sign of r_{pb}). This was done, and the results are reported in row 3 of Table 1. The degree to which individual performance is associated with the SDSVs now appears to be considerably higher for most scales.

The correlations reported in rows 1 and 3 of Table 1 are limited by the reliabilities of the correlated measures; they were therefore corrected for attenuation. Corrected algebraic and modular mean r_{pb} are shown in rows 4 and 5 of Table 1. The reliability

Table 1. *Summary of correlational analyses for group I* ($n = 50$) *and group II* ($n = 12$)

Row	Group	Correlation type	Correlation between	Notes	*L*	*F*	*K*	*Hs*	*D*	*Hy*	*Pd*	*Mf*	*Pa*	*Pt*	*Sc*	*Ma*	*SD*
1	I	r_{pb}	Self-descr. and SDSV	Algebraic mean	0·13	0·72	0·02	0·09	0·09	0·04	0·17	0·14	0·37	0·14	0·45	0·11	0·27
2	II	r_{pb}	Self-descr. and SDSV	Algebraic mean	0·14	0·69	−0·07	0·15	0·09	0·03	0·14	0·12	0·38	0·07	0·36	0·11	0·23
3	I	r_{pb}	Self-descr. and SDSV	Modular mean	0·25	0·72	0·19	0·37	0·29	0·25	0·25	0·22	0·37	0·29	0·45	0·19	0·34
4	I	r_{pb}	Self-descr. and SDSV	Algebraic mean corr. for attn.	0·16	0·83	0·03	0·11	0·11	0·05	0·21	0·17	0·48	0·19	0·53	0·15	0·35
5	I	r_{pb}	Self-descr. and SDSV	Modular mean, corr. for attn.	0·31	0·83	0·26	0·45	0·37	0·32	0·32	0·26	0·49	0·39	0·53	0·25	0·44
6	II	r_ϕ	Self-descr. Test/Retest	Algebraic mean	0·80	0 80	0·59	0·73	0·66	0·66	0·67	0·72	0·63	0·60	0·73	0·61	0·63
7	—	r_{pb}	All items answered in SD dir. and SDSV	Max. r_{pb} possible	0·80	0·97	0·88	0·98	0·91	0·90	0·89	0·87	0·83	0·91	0·95	0·80	0·95
8	I	r	SD scale and MMPI scales	With item overlap	0·25	−0·66	0·67	−0·66	−0·70	−0·33	−0·47	−0·35	−0·47	−0·89	−0·87	−0·27	—
9	I	r	SD scale and MMPI scales	Without item overlap	0·25	−0·58	0·54	−0·63	−0·64	−0·16	−0·41	−0·35	−0·38	−0·86	−0·85	−0·37	—
10	I	r	r_{pb} and raw score	—	0·34	−0·99	0·48	−0·97	−0·85	−0·63	−0·76	−0·47	−0·61	−0·94	−0·93	−0·17	0·92
11	I	r	SD scale and MMPI scales	With overlap corr. for attn.	0·35	−0·80	0·91	−0·78	−0·85	−0·38	−0·55	−0·48	−0·59	−1·00*	−1·00*	−0·39	—
12	I	r	SD scale and MMPI scales	Without overlap corr. for attn.	0·35	−0·84	0·85	−0·82	−0·89	−0·20	−0·54	−0·48	−0·50	−1·00*	−1·00*	−0·59	—
13	I	r	r_{pb} on *SD* Scale and r_{pb} on MMPI scales	—	0·27	0·60	0·49	0·64	0·82	0·78	0·50	0·59	0·51	0·80	0·78	0·47	—

* Corrected $r > 1{\cdot}00$.

Table 2. *Variability of SD measures* (r_{pb}) *for each MMPI scale* ($n = 50$)

MMPI scale	$s_{z'}$*	Limits of range	% r_{pb} significantly less than zero ($P < 0.05$)	% r_{pb} significantly greater than zero ($P < 0.05$)
L	0·288	−0·41 to 0·66	0	8
F	0·381	0·31 to 0·97	0	100
K	0·272	−0·40 to 0·84	4	4
Hs	0·529	−0·81 to 0·90	18	20
D	0·351	−0·50 to 0·69	16	30
Hy	0·335	−0·65 to 0·71	20	26
Pd	0·286	−0·52 to 0·65	4	30
Mf	0·183	−0·30 to 0·50	2	24
Pa	0·269	−0·12 to 0·89	0	54
Pt	0·368	−0·50 to 0·77	14	28
Sc	0·336	0·00 to 0·87	0	70
Ma	0·221	−0·27 to 0·55	0	18
SD	0·387	−0·37 to 0·91	2	38

* s.d. of r_{pb} expressed in terms of z'.

of the SDSVs, obtained from student responses, was taken at 0·97 as reported by Boe & Kogan (1963) over 100 items. This reliability coefficient was corrected for the item length of each MMPI scale by the Spearman–Brown prophecy formula before the corrections for attenuation were made. The reliability of individual item responses was determined from the test–retest data of the twelve subjects of group II. For each subject on each MMPI scale, the test–retest reliability of dichotomous true/false responses to individual items was computed by phi correlation coefficients (r_{ϕ}). The mean r_{ϕ} for each MMPI scale is given in row 6 of Table 1. (The appropriateness of correcting a mean correlation for attenuation with a mean reliability coefficient was considered. Individual correlations might first be corrected for attenuation by individual reliability coefficients, and then the mean corrected correlation taken. Both procedures were possible, and were followed, for subjects in group II. The latter procedure yielded slightly higher mean corrected correlations. The rank order correlation between the mean r_{pb} corrected by the two procedures across the thirteen MMPI scales was 0·97.)

It can be argued that another influence will tend to restrict the size of the mean r_{pb}: r_{pb} has an upper limit which is less than one. When the continuous variable is normally distributed and there is a 50–50 % cut in the dichotomous variable, the maximum r_{pb} is 0·80, and, as McNemar (1962) has stated: 'These two coefficients [r_{pb} and r_b] are not on the same scale . . . Even if the dichotomous variable were a genuine point variable, r_{pb} as an expression of the degree of relationship would not be comparable either to r_b or to the product-moment r between the two variables measured in a graduated fashion' (p. 193). The distributions of the SDSVs (the continuous variable) for the MMPI scales used in this study do not approach normality; rather are they characteristically bimodal, with modes on each side of neutral desirability. When all items with SD scale values are answered 'true', and items with SUD scale values are answered 'false', this condition tends to increase the maximum size of the r_{pb} to the values reported in row 7 of Table 1. Variability in the maximum size of r_{pb} raised the possibility that variability in mean r_{pb} might be accounted for in terms of this statistical artifact. In order to evaluate the importance of this statistical consideration, the following three variables were defined and rank order correlations (ρ) between them were computed across the thirteen MMPI scales: (*a*) modular mean r_{pb} for group I; (*b*) the average difference between the SDSV of items in each MMPI scale and neutral desirability (this variable is subsequently referred to as the SD intensity variable); and (*c*) maximum r_{pb} on each MMPI scale. The correlations among these variables were: $\rho_{ab} = 0{\cdot}94$; $\rho_{ac} = 0{\cdot}71$; $\rho_{bc} = 0{\cdot}79$. Thus, statistically significant relations were obtained between mean r_{pb} and variables (*b*) and (*c*). Partial correlations were then calculated in an effort to evaluate more precisely the influence of these two variables on mean r_{pb}. They were: $\rho_{ab.c} = 0{\cdot}88$; $\rho_{ac.b} = -0{\cdot}16$. From the foregoing analysis it is apparent that variability in mean r_{pb} across MMPI scales can be largely attributed to the SD intensity variable, and not to differences in the upper limit of r_{pb}.

Correlations with the SD response tendency

Correlations between the *SD* scale raw scores and the MMPI scale raw scores (without the *K* correction), which were computed across the fifty subjects of group I, are shown in row 8 of Table 1. The size of these correlations is somewhat inflated by item overlap between the *SD* scale and the MMPI scales. Consequently, the MMPI scales were rescored after overlapping items had been eliminated, and the resulting thirty-nine item *SD* scale was correlated with the corrected raw scores of the MMPI scales. The no-overlap correlations are presented in row 9 of Table 1. The elimination of overlapping items reduced the size of the correlation for all scales, except the *Ma* scale. For the *Ma* scale, the overlap tended to produce a positive correlation, although the correlation with the *SD* scale was negative. Thus, when the overlap was eliminated, a larger negative correlation resulted.

The second measure of individual differences between subjects in the SD response tendency was r_{pb}. The r_{pb} for a specific scale were also correlated with the raw scores on that scale (see row 10 of Table 1) just as scores on the *SD* scale were. However, since the r_{pb} and the raw scores were derived from the same sample of items, it is not surprising that their correlations are generally higher than those with the SD scale which came from a different sample of items. The size of these correlations should vary with the proportion of items in each scale which are keyed for SD responses. If all items are keyed in the SD direction, as they are in the *SD* scale, the raw score count and r_{pb} are simply alternative ways of obtaining an SD score. Boe & Kogan (1964) have reported correlations of 0·95 and 0·97 in two different subject samples between these two methods. If all items are keyed in the SUD direction, as they are in the *Hs* and *Pt* scales, the correlation should be near $-1{\cdot}00$; and as the proportion approaches 50% the correlations should be near zero. The actual relation, as measured by rank order correlation, ρ, between the correlations in row 10 of Table 1 and the proportion of items keyed for SD is 0·94.

The magnitude of the correlations between raw scores on the MMPI scales and the *SD* scale are also limited by their reliabilities. Therefore, they were corrected for attenuation by using Rosen's (1953) reliability coefficients of scores on MMPI scales and Edwards's (1957) reliability of 0·83 for the *SD* scale. These corrected correlations are presented in row 11 of Table 1. The reliability coefficients for the shortened no-overlap MMPI scales were computed from Rosen's (1953) reliability coefficients by the Spearman–Brown prophecy formula. The corrected no-overlap correlations of the MMPI scales with the *SD* scale are given in row 12 of Table 1. A comparison of rows 11 and 12 reveals little difference between the overlap and no-overlap correlations with the *SD* scale when both are corrected for attenuation.

The final analysis concerned the similarity of SD response tendency measures. If the *SD* scale is a useful general measure of the SD response tendency, as Edwards (1957) intended, when r_{pb} are computed from it they ought to be highly correlated with r_{pb} computed from the items of the other MMPI scales. These correlations (r) are reported in row 13 of Table 1. Half of these correlations are less than 0·60, and none exceeds 0·82. Since each r_{pb} is an SD score, the coefficients in row 13 of Table 1 are further evidence that measures of the SD response tendency vary as a function of item content (Boe & Kogan, 1964; Crowne & Marlow, 1960; Schultz, Kogan & Chapman, 1962; Wiggins, 1959). Correlations of MMPI scales with the *SD* scale, therefore, should not be expected to estimate accurately the amount of between individual variance associated with the SD response tendency for all MMPI scales because, as content differs, so does the SD responses tendency.

DISCUSSION

The most meaningful measure of the average amount of within individual variability in self-descriptions that is associated with SDSVs for specific MMPI scales is the modular mean r_{pb} corrected for attenuation (row 5, Table 1). With the exception of the *F* scale, these mean correlations are considerably lower than uncorrected correlations typically found between the probability of endorsement and SDSVs; once again they demonstrate that, on the average, only a minor portion of individual performance is associated with SDSVs. Considering all thirteen scales, the median amount of within individual variability accounted for by SDSVs is 14 % in a sample of psychiatric patients. Viewed in this way, individual MMPI performance is certainly not dominated by SD response tendencies.

Since individual item responses were correlated with mean SDSVs, it is possible that the relatively low mean r_{pb} would be higher if an individual's item responses had been correlated with *his own* judgements of SDSV instead of with the mean SDSV of a group of subjects. Whether such correlations with *individual* SDSVs are higher than with *group* SDSVs seems to depend upon the subjects used. Taylor's (1959) results (which are most comparable to the present data because he also used psychiatric patients as subjects) indicate that the difference is not significant (the individual correlations were actually a little lower on the average). For a group of college students, however, Boe & Kogan (1963) found that using individual SDSVs instead of group SDSVs significantly raised the r_{pb} from a mean of 0·32 to a mean of 0·54 when corrected for attenuation. Little more can be said until further work on this problem is accomplished.

Another obvious characteristic of the corrected mean r_{pb} reported in row 5 of Table 2 is their variability across MMPI scales. The partial rank correlation of 0·88 between arithmetic mean r_{pb} and the SD intensity variable accounts for most of this variability. Most *F* scale items (corrected mean $r_{pb} = 0{\cdot}83$), for example, have either extreme SD or SUD scale values (in nearly equal proportion), while the distribution of the SD scale values of *Mf* scale items (corrected mean $r_{pb} = 0{\cdot}25$) is approximately rectangular. Thus, as the variability of the SDSVs of a sample of items decreases, so does the SD response tendency as measured by mean r_{pb}. An implication of findings such as these for attempts to develop SD-free scales is that one of several suitable approaches would be to select items with very similar SDSVs.

Individual r_{pb} for a particular MMPI scale can also be used as scores which measure between individual differences in the SD response tendency for those particular items. The variability measures reported in Table 2 indicate that there are wide individual differences in this respect on most scales, and the percentage of significant r_{pb} shows that almost as much SUD response tendency (negative r_{pb}) appears on scales of the 'neurotic triad' (*Hs*, *D*, *Hy*) as SD response tendency (positive r_{pb}). The algebraic mean r_{pb} of these three scales are only slightly above zero. In contrast with the three scales of the 'neurotic triad', there is relatively little SUD response tendency in the three scales of the psychotic group (*Pa*, *Pt*, *Sc*). Apparently, the patients in this sample are more willing to acknowledge traits (which usually have SUD scale values) related to neuroticism than to psychoticism. In a psychiatrically normal population, Boe & Kogan (1963) found little evidence of SUD response tendencies, i.e. of negative r_{pb}.

The best estimate of between individual variability in raw scores for specific MMPI scales which is associated with the SD response tendency is the correlation between raw scores and SD scores (r_{pb}) computed from the same items (row 10, Table 1). The rank correlation across the thirteen scales of 0·94 between these correlations and the proportion of items keyed for SD response accounts for most of their variability. At the limiting proportions (0 and 1), the raw score counts and r_{pb} are almost perfectly related, and, consequently, almost completely confounded. They are simultaneously two different methods for computing (*a*) SD scores *and* (*b*) scores for the behaviour which the scales are presumed to measure. From these and similar findings, it is clear that a second method for developing SD-free scales is to select a sample of items in which (when keyed for the behaviour being measured) the proportion keyed for SD approximates 0·50.

In comparing (*a*) the most meaningful measure of average *within* individual variability in self-descriptions associated with SDSVs (modular mean r_{pb}) with (*b*) the best measure of *between* individual variability in raw scores associated with the SD response tendency (correlation between raw scores and r_{pb} for each scale), it is apparent that the latter group of correlations are much higher, even when reliability considerations are taken into account. If the same question is put inappropriately to both methods (for instance, to what extent is MMPI performance associated with SD?), quite different answers will be forthcoming. However, the fundamental differences in the types of variable being correlated would suggest that they are complementary rather than comparable, and that each provides information about unrelated aspects of MMPI performance: the correlation between them across the thirteen scales is $\rho = 0{\cdot}81$ ($P < 0{\cdot}01$), but when the SD intensity variable is partialled out, the partial coefficient is only $\rho = -0{\cdot}05$.

Regardless of whether the SD response tendency is eventually identified as a general personality trait, as SD faking, or as a combination of both, significant advantages accrue from removing it as a confounding variable from measures of other traits. Purer and more meaningful measures of hostility and extraversion, for example, result if they are not confounded with the tendency to give socially desirable responses in self-description. Considering the SD response tendency as a personality trait in its own right, it is important that it should also be measured separately from other traits. Two approaches to unconfounding these measures that come out of the present results have been discussed.

References

Block, J. (1965). *The Challenge of Response Sets.* New York: Appleton-Century-Crofts.

Boe, E. E. & Kogan, W. S. (1963). Social desirability response set in the individual subject. *J. consult. Psychol.* **27**, 369.

Boe, E. E. & Kogan, W. S. (1964). An analysis of various methods for deriving the social desirability score. *Psychol. Rep.* **14**, 23–9.

Crowne, D. P. & Marlow, D. (1960). A new scale of social desirability independent of psychopathology. *J. consult. Psychol.* **24**, 349–54.

Dahlstrom, W. G. & Welsh, G. S. (1960). *An MMPI Handbook: a Guide to Use in Clinical Practice and Research.* Minneapolis: University of Minnesota Press.

Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.* **37**, 90–3.

Edwards, A. L. (1957). *The Social Desirability Variable in Personality Assessment and Research.* New York: Dryden.

Edwards, A. L. (1961). Social desirability or acquiescence in the MMPI? A case study with the *SD* scale. *J. abnorm. soc. Psychol.* **63**, 351–9.

Edwards, A. L. (1963). A factor analysis of experimental social desirability and response set scales. *J. appl. Psychol.* **47**, 308–16.

Edwards, A. L. & Diers, C. J. (1962). Social desirability and the factorial interpretation of the MMPI. *Educ. psychol. Measmt* **22**, 501–9.

Edwards, A. L., Diers, C. J. & Walker, J. N. (1962). Response sets and factor loadings on sixty-one personality scales. *J. appl. Psychol.* **46**, 220–5.

Edwards, A. L. & Heathers, L. B. (1962). The first factor of the MMPI: Social desirability or ego strength? *J. consult. Psychol.* **26**, 99–100.

Edwards, A. L. & Walker, J. N. (1961). Social desirability and agreement response set. *J. abnorm. soc. Psychol.* **62**, 180–183.

Edwards, A. L. & Walsh, J. A. (1963). The relationship between the intensity of the social desirability keying of a scale and the correlation of the scale with Edwards' SD scale and the first factor loading of the scale. *J. clin. Psychol.* **19**, 200–3.

Hanley, C. (1956). Social desirability and response bias in the MMPI. *J. consult. Psychol.* **25**, 13–20.

McNemar, Q. (1962). *Psychological Statistics*, 3rd ed. New York: Wiley.

Rosen, A. (1953). Test–retest stability of MMPI scales for a psychiatric population. *J. consult. Psychol.* **17**, 217–21.

Rosen, E. (1956). Self-appraisal, personal desirability, and perceived social desirability of personality traits. *J. abnorm. soc. Psychol.* **52**, 151–8.

Schultz, C. B., Kogan, W. S. & Chapman, H. (1962). Favorability, unfavorability and content considerations in SD scales. *Psychol. Rep.* **10**, 619–22.

Taylor, J. B. (1959). Social desirability and MMPI performance: The individual case. *J. consult. Psychol.* **23**, 514–7.

Wiggins, J. S. (1959). Interrelationship among MMPI measures of dissimulation under standard and social desirability instructions. *J. consult. Psychol.* **23**, 419–27.

Wiggins, J. S. & Rumrill, C. (1959). Social desirability in the MMPI and Welsh's factor scales A and R. *J. consult. Psychol.* **23**, 100–6.

(*Manuscript received* 14 *July* 1965)

Brit. J. Psychol (1966), 57, 1 *and* 2, *pp.* 171–185
Printed in Great Britain

THE BROOK REACTION TEST OF INTERESTS

By A. W. HEIM and K. P. WATTS
The Psychological Laboratory, Cambridge

A test of interests is described which is open-ended yet objectively scorable and which yields validatory evidence for 20 of the 22 relevant interests. There is some suggestion too that the test may be useful in estimating degree of stability and social adjustment. A number of experiments have been conducted, using as subjects university and other students, naval ratings, secretaries and graduates. The test appears to be suitable for such subjects and also for senior school-children.

Interests testing, if valid, may have value both theoretically and practically. In theoretical psychology the results may be used, for instance, in developing personality theory, in examining the relations between interests and abilities, and in furthering research in human motivation and learning. On the practical side a valid interests test may be useful in vocational guidance and selection, in clinical work and in educational guidance such as helping waverers and all-rounders to decide in what to specialize, whether at school or at university level.

The *Strong Vocational Interest Blank* (Strong, 1959*a*,*b*), the *Kuder Preference Record* (Kuder, 1960), the *Study of Values* (Allport, Vernon & Lindzey, 1960) and many other questionaires, purport to measure interests and to yield a 'profile' of the subject tested. Such questionnaires, however, are open to a number of criticisms. They are of the multiple-choice, forced-choice variety. Thus the subject has to select which of several modes of activity or ways of life he would most prefer, which he would like next best, etc., but he is not allowed say that they all appeal to him equally—nor that they all nauseate him. This authoritarianism is frustrating for the serious-minded subject; it does not worry the light-hearted and flippant, who may find in such tests a welcome opportunity for leg-pulling.

Moreover, some of the items in these questionnaires are disconcertingly double-barrelled. In the *Study of Values* we find, for example, (i) 'Because of the aggressive and self-assertive nature of man, the abolition of war is an illusory ideal. (*a*) Yes. (*b*) No.', and (ii) 'You are asked to wait for a few minutes in a strange living room; are you more likely to judge (*a*) the owner's knowledge and intellectuality as shown by the books in his book cases: (*b*) his friendliness and hospitality as shown by the photographs on the walls, and the general comforts of the room?' The subject who feels with regard to (i), for instance, that 'the abolition of war is an illusory ideal' but *for economic reasons* or, with regard to (ii), that one cannot judge 'friendliness and hospitality' on the basis of photographs and general comforts, is rendered inarticulate and sometimes furiously unco-operative at having to place himself, as he feels, in a false position. There is no way of distinguishing between the critical or unconforming and the waggish or phlegmatic, in terms of their responses. To ignore such difficulties and defend such tests in virtue of their (often unconvincing) statistical validation is not good enough.

It is not only the immediate reactions of the subject which constitute objections to this type of test—although, in the view of the writers, these reactions are far more important than is evidently believed by most psychometrists. Such tests proliferate further psychological and statistical difficulties. They tend to provoke 'acquiescence response style' and 'desirability response style' answers to the test questions (see, for example, Messick & Ross, 1962). These terms signify merely the tendency of subjects, other things being equal, to reply in the affirmative rather than the negative and to try to give as good an impression of themselves as possible. Such tendencies have long been recognized outside the psychological laboratory, and they have now been experimentally demonstrated. The consequence has not been to abandon the traditional questionnaire and seek a new approach, but to build in 'stylistic measures' and 'lie detectors'.

Tiresome response styles are liable to occur, even when other things are equal. More often than not, however, questionnaires are administered with a purpose—such as selection, for instance—and in such situations the subject's natural tendency to answer in a 'desirable' manner is exacerbated. Even Eysenck, a protagonist of questionnaires, has said that 'where people are motivated to try and give as good an account of themselves as possible, questionnaires are almost entirely useless'. (Eysenck, 1964). Yet the number of situations in which the subject is *not* so motivated, whilst nonetheless being motivated to co-operate with the requisite willingness and honesty, must be exceedingly small.

A further objection to these questionnaires is that a high score on one interest automatically ensures a low score on one or more other interests, and conversely. Such tests imply the assumption that there is so much 'interestability' in all of us—an *equal* amount in all of us—and therefore if we have many interests (or no interests) our profile will be flattish, i.e. a rather uniform and low level of enthusiasm generally will emerge. Again someone lacking markedly in interest x will be bound to show up highly on interest y or z, as a direct result of the scoring system and the forced-choice technique. Yet the existence of some people whose interests are both wide and deep (and of others whose interests are both narrow and shallow) is observable.

Finally, the multiple-choice technique, by its very nature, does not allow the subject to show any enterprise or individuality in his interests. No scope is offered for someone with unusual interests, or even with strong contra-suggestibility, to portray himself, whether he will or no. These tests which claim to be assessing an important facet of personality are 'closed-ended' to the highest degree. This is surely less permissible in orectic tests than in cognitive tests, where there is intended to be one, and only one, right answer.

Thus our aims in devising the Brook Reaction test were to develop a test of interests which is open-ended and yet objectively scorable and which offers a very wide spectrum of interests; a test in which depth and width of interest are compatible (as are superficiality and narrowness); which contains no forced choices and no opportunities for manifesting 'acquiescence' or 'desirability response styles'; a test whose results will be unaffected by the situation in which it is taken; and which allows scope for idiosyncrasies to display themselves. Finally, though far from least in importance, the test must correlate highly with other measures of interests.

THE TEST

Historical

The Brook Reaction test is a form of word-association test (Jung, 1918; Kent & Rosanoff, 1910, 1911; Vernon, 1953). It derives its name from the late David Brook, who was largely responsible for selecting the original stimulus words and for testing subjects when the test was in its earliest form (Brook & Heim, 1960). Initially, the test consisted of 106 ambiguous words, such as CANVAS(S), TRAIN and SERVICE. The words were presented orally and individually to the subject, who was instructed to reply with the first word that entered his head. To CANVAS he might respond, for instance, 'sailing' or 'painting' or 'tory', and so on. Most of the ideas aroused by the stimulus words were related to one of nine different interests, such as *outdoor* (illustrated by 'sailing' above), *aesthetic* (illustrated by 'painting', above), *political* (illustrated by 'tory' above). The other six interests were *sport*, *religious*, *scientific*, *practical*, *academic* and *social*.

Validation was attempted by means of a structured interview, incorporating a short factual list of questions. On the basis of their interview answers, subjects were rated for each interest on a 4-point scale and these ratings were correlated with their (weighted) test score. Using Kendall's *tau*, six of the interests yielded correlations significant with $P < 0{\cdot}01$ and one yielded a correlation significant with $P < 0{\cdot}05$. This was regarded as sufficiently promising to develop the test further.

Stimulus words and their responses

Since 1960, many changes have been made. The 106 stimulus words have been reduced to eighty, because some of the original words were found to be unsatisfactory or redundant. These unsatisfactory stimuli either produced a high proportion of 'unclassifiable' responses, i.e. responses which failed to indicate any clear interest of any kind (e.g. FAST); or they were found to hold too limited and topical an interest (e.g. EDEN); or they appeared to create embarrassment for some subjects (e.g. VICE, PICK-UP).

A further change is the classification and inclusion of responses which were made to a *misheard* stimulus word (e.g. hearing PLOT as CLOT, or SEW as SOUL), on the grounds that there may well be relevant psychological reasons for the particular mishearing. The eighty stimulus words are always presented in the same order, since it has been established that mental set does not arise as a result of the order in which the stimuli are heard (Hargreaves, Heim & Watts, 1963). A representative sample of ten of the present stimulus words is given in Table 6 (p. 185).

The interests covered

Although the number of stimulus words has been reduced to eighty, the number of interests has increased to twenty-two. A brief description of these interests is given in Table 1. (This is in fact the rating sheet which has been given to subjects, for validation purposes, after they have taken the test.) Not only has the number of categories more than doubled, but some of the original interests now have a different emphasis. For example, *aesthetic* used to include *literary* but it has been found more satisfactory to separate interest in literature from that in music and the visual arts.

The *outdoor* category used to be very wide and vague; we have now separated *agricultural* interests, e.g. 'hay' in response to Bail (Bale), 'wheat' in response to Sew (Sow), from non-agricultural outdoor responses (such as 'mountain' in response to Pass). Again, we had initially used the term *social* to cover interest in people generally. We found it necessary, however, to split this into three categories, one designating interest in *people* (adults and children) as individuals (*Pe*); one designating interest in groups, primarily from a *humanitarian* point of view (*H*); and one designating sociability—taking *pleasure in social functions* as such (*D*). We found it desirable, too, to split our original *scientific* category into two, *biological* and *physical*, despite the fact that these two categories have many concepts and many terms in common.

Table 1. *The rating sheet used in validating the Brook Reaction test, including brief descriptions of the twenty-two interests*

Rating on.. by ....

Please read the instructions carefully and then fill in the appropriate number in each of the 22 spaces.

Instructions

5 = Extremely interested.
4 = Very much interested.
3 = Some definite interest.
2 = Mildly interested.
1 = Minimum of interest. Very rarely do anything about it.
0 = No interest at all. Never think, talk or do anything along these lines.
−1 = Positively anti. Actively object to the topic in question.

Interests

A = aesthetic. Interests in the arts, especially music, painting and sculpture ....
B = business. Finance, trade, commerce, economics, money-making ....
C = clothing. Personal appearance and dress. Perhaps in materials (textiles) ....
D = dances, parties and social functions generally ....
E = entertainment. Theatre, films, wireless, television ....
F = food and drink ....
G = agriculture and farming; gardening and horticulture ....
H = humanitarian. Interested in people generally, rather than in individual persons. Social service, welfare, organizations ....
I = intellectual interests. Discussing and thinking about abstract ideas: includes historical and theoretical matters ....
K = practical. 'Do it yourself' interest; keenness on crafts, gadgets, machines; helping at home ....
Li = literary. Books and writers; prose and poetry ....
Lw = law. Legal matters and things connected with the law generally ....
M = military, including all the Armed Forces ....
O = outdoor activities. Camping, boating, fishing, walking, mountaineering. Love of the countryside ....
Pe = people. Interested in people as individuals; in children; in human personality ....
Pl = political. Current affairs, elections, political parties, unions ....
R = religion. Church-going, the Bible; and/or religion plays an important part in life ....
SB = biological sciences: medical, botanical, zoological interests ....
SN = physical sciences: physics, geology, mathematics ....
Sl = secretarial. Interest in office and clerical matters. Filing papers, organizing, correspondence ....
Sp = sport. Active participation in sports and athletics, whether outdoor (e.g. cricket) or indoor (e.g. boxing) ....
T = travel. Visiting new places, including going abroad. Geographical interests ....
X = any *other* specially keen interest. (Say what it is)

Some of the additional interests have been added, unwittingly, by the subjects themselves: a substantial minority of them responded to a meaning of a stimulus word, previously unrecognized by us. For example, 'army, horse' in response to GRAZE (GREYS) suggested the addition of *military* interests. Other new interests have come into the test because they were fairly frequently mentioned at the bottom of our early rating sheet as an 'other specially keen interest'. These later categories comprise: *business, clothing and appearance, entertainment, food and drink, legal, secretarial* and *travel*. Thus we are able to classify satisfactorily many responses which would have fitted uncomfortably, or not at all, into the older categories.

Test procedure

The test has been transformed into a paper and pencil test, thus enabling it to be given to groups; and instead of giving one-word responses, the subject now gives as many responses as he can write in the 12 sec time limit allowed. A tape has been made, recording the instructions, the preliminary examples and the eighty test words, the latter separated by 12 sec intervals. It takes about 20 min. Most of the data reported in this paper are based on testing with the tape. The test instructions are as follows: 'This test is a Reaction test. It is very simple and there are no right or wrong answers. I am going to say a word and I want you to write down the first word that comes into your head after that, and then what the word you have written reminds you of, and so on.' Two preliminary examples are then given: HOT—'cold, ice, skating and HOT—'sun, rise, east'. The subjects are then given 12 sec in which to practise on the (equally unambiguous) word, WINDOW. It is thought that these preliminary examples of words with only one meaning helps to establish a set towards one-meaning interpretation of the test words—though, as anticipated, the single meaning taken in the test proper varies from subject to subject. A short rest of about 15 sec is given after the first forty words.

Thus the subjects are not told the purpose of the test beforehand; they do not suspect it during the course of the test; and very few of them realize that all the stimulus words are ambiguous. Most manage to write at least two responses in the time allowed; many produce three; and responses of four, and even five words are not unknown. Whilst a few subjects respond to several different meanings of the stimulus word, the majority confine themselves to one theme per stimulus word—though some people do go off at a tangent to one of their own responses, e.g. stimulus ANGLE eliciting the response 'acute, chronic'.

The system of scoring

In the earlier form of the test the subject was asked for a one-word answer to each stimulus word. Thus only one interest category could be represented in any one response. The scores were weighted, however, the argument being that someone who responded to a rare meaning of the stimulus word (e.g. 'novelist' to FIELDING) was more deeply interested in *literature* than a subject who replied with 'cricket'—the most popular type of response to this stimulus.

Four changes have been made in the scoring system. First, weighting has been abolished, since there was no significant difference between the correlations of the ratings with (i) weighted scores and (ii) unweighted scores. The establishment of norms now takes care of this particular problem. Secondly, a total of four points is scored for each response. Thus, since there are eighty stimulus words, 320 points are available for each subject. Thirdly, a subject who responds, say, to two different meanings of the stimulus word divides his four points between these two meanings. For instance, someone who wrote 'theatre, Apostles' in reply to ACT would score two for *entertainment* and two for *religion*. Fourthly, we can now allow for the fact that certain categories, and certain responses, may span more than one interest. For example, *physical science* and *practical* interests may overlap (as in the response 'battery' to stimulus SELL (CELL)); *intellectual* and *legal* may overlap (as in the response 'justice' to SCALES).

The present system of scoring allows the four points available to be split, in order to allow for such overlapping and also for occasional uncertainty on the part of the scorer. The element of uncertainty has, however, been greatly reduced by the system of encouraging people to write more than one word. As a result, responses which used to be hard to classify, such as 'wood' in reply to CARVING (CALVING), can now be scored largely as *aesthetic* if followed by 'figure', and as *practical* if followed by 'saw'. Despite this technique, there are still certain responses which do not appear to fit neatly into any one, or any combination, of the twenty-two interests. These are designated as *unclassifiable* (*Un*). An example would be 'corner' in reply to ANGLE.

A subject who responds with one word only is regarded as having omitted part of his answer and this is recorded as *Mo*. Thus, for instance, someone who gave 'music' alone, in reply to the stimulus word KEY (QUAY), would score *A* 2, *Mo* 2. Whereas if he gave 'door', alone, he would score *Un* 2, *Mo* 2. Each *Mo* carries two points: thus a complete blank is scored *Mo* 4. Certain responses (which may or may not be otherwise classifiable) we have designated as ***sinister*** (*Sin*) following Rorschach terminology (Alcock, 1963). These are responses such as 'drowning' or 'death' or 'blood'—when such responses do not appear relevant. For example 'blood' in reply to GRAZE or VEIN (VAIN) is not categorized as *Sin* Whereas 'blood' as first response in reply to the stimulus word ACT is so categorized. A record is kept of the number of times a subject (disregarding the instructions) responds to two or more distinct meanings of the stimulus word. These are termed 'double (or triple) theme responses'. An example of this would be the reply 'rock, sausage' to the word ROLL (scoring *D* 2, *F* 2).

AIMS AND METHODS

The main aims of the experiments described in this paper were to establish norms and to obtain validatory evidence for the Brook Reaction test. The groups tested include university students reading natural science and a variety of other subjects, including architecture and the humanities; naval ratings; some Royal Naval College engineering students; a small group of women secretaries; and a small group of graduate executives in an investment research organization. (A number of schoolboys and schoolgirls have also been tested but their results are not discussed here.)

The validation criteria are of three kinds: (*a*) ratings, (*b*) scores gained on other tests, (*c*) criterion groups. The rating sheet (see Table 1) has been completed by all the subjects, i.e. they rated themselves on each interest, after having taken the test. So far, the test has been given, and the ratings made, in the experimental situation only. In addition to these self-ratings, most subjects have got at least one other person to rate them, independently, usually a friend or a relative. The self-rating has always been given a weighting of 50 %, the other one or more ratings making up the remaining 50 %. The agreement between self-ratings and the other ratings was found to be satisfactorily high, differences of more than one point being rare, those of more than two points being exceptional.

Judging from the frequency distributions of the ratings, some of the interests are implicitly considered to be more desirable than others. *Travel*, for instance, is rarely self-rated as lower than 3; *secretarial* and *legal* are rarely self-rated more highly than 3; the rating of other interest categories, such as *business* and *military*, depend largely on the context. Students at a Royal Naval Engineering College, for instance, commented in informal discussion that they could not very well rate themselves or their fellows as lower than 3 on interest in *military* matters. The majority of the interests, however, yielded a wide range of ratings: all the way from +5 to 0 (and sometimes to −1).

THE EXPERIMENTS

The experiments are outlined under three heads: *A*, Cambridge undergraduates reading experimental psychology as one of several subjects for Part I of the Natural Science tripos; *B*, naval ratings; *C*, criterion groups, including university students reading a variety of subjects and belonging to a variety of clubs and societies, R.N. engineering students, women secretaries, and graduates engaged in investment research.

A. Cambridge psychology students

Procedure

Ninety-three students reading for Part I of the Natural Science tripos were given the test. In addition to psychology, they were reading two or three other subjects. These were extremely varied, including, for instance, anatomy, biochemistry, botany, chemistry, crystallography, history and philosophy of science, geology, mineralogy and petrology, mathematics, pathology, physics, physiology and zoology. They took this test, the *Study of Values* and a test of introversion–extraversion (Neymann & Kohlstedt, 1929) as part of their practical course.

After having taken the Brook test, they all completed the rating sheet (see Table 1) and the majority got themselves rated, in addition, by at least one other person who, in their view, knew them sufficiently well. The numbers were as follows:

		n
Brook Reaction and *Study of Values*:	with two or more ratings	59
	with only one rating	21
Brook Reaction (no *Study of Values*):	with two ratings	11
	with only one rating	2

Results

Table 2 shows the rank correlations (Spearman's rho; Siegal, 1956) between Brook Reaction score and mean rating. Only those seventy students who had two or more ratings were used for these correlations.

In this paper, ratings of -1 have been omitted, hence the variation in *n*. It was felt that people who feel strongly enough about *religion*, for instance, to rate themselves as -1 were probably militant atheists and, as such, have in a sense a greater interest in religion than those who give mild ratings of 0 or 1. In practice, ratings of -1 were almost entirely confined to *religion*, *military*, *secretarial* and *sport*.

Table 2. *Comparisons between Brook scores and mean interest ratings for psychology students* ($n_{\text{max.}} = 70$)

Interest rating							Brook score	
Mean	Range	Interest category	*n*	ρ	*t*	$P <$ (one-tailed)	Mean	Range
3·43	½–5	*A*	70	0·322	2·803	0·005	13·31	3–33
2·02	0–5	*B*	63	0·125	0·984	N.S.	14·14	4–27
2·73	½–4½	*C*	69	0·402	3·594	0·0005	9·25	0–22
3·08	½–5	*D*	65	0·374	3·201	0·005	4·83	0–11
3·31	½–5	*E*	70	0·061	—	N.S.	10·21	0–32
3·05	1½–5	*F*	70	0·067	—	N.S.	17·56	1–34
1·25	0–5	*G*	70	0·218	1·842	0·05	7·21	0–18
2·42	0–4½	*H*	68	0·252	2·126	0·025	6·71	0–18
3·27	½–5	*I*	70	0·055	—	N.S.	10·73	1–20
2·40	½–4½	*K*	68	0·408	3·630	0·0005	15·56	1–32
3·01	½–5	*Lt*	70	0·289	2·490	0·01	6·41	0–16
1·03	0–2½	*Lw*	66	0·074	—	N.S.	2·86	0–6
0·97	0–5	*M*	45	0·007	—	N.S.	0·78	0–7
3·26	0–5	*O*	69	0·314	2·707	0·005	29·04	12–60
3·62	1–5	*Pe*	70	−0·042	—	N.S.	31·67	16–59
2·59	1–5	*Pl*	63	0·355	2·965	0·005	8·59	0–19
2·50	0–5	*R*	57	0·180	1·356	0·1	8·89	2–25
3·07	0–5	*SB*	68	0·389	3·430	0·001	10·56	0–32
2·71	0–5	*SN*	68	0·520	4·945	0·0005	12·5	1–34
1·25	0–4	*Sl*	60	0·127	0·974	N.S.	6·27	2–12
2·77	0–5	*Sp*	57	0·455	3·789	0·0005	16·79	1–43
3·67	1–5	*T*	70	0·262	2·239	0·025	10·51	2–29

For thirteen of the twenty-two interests, positive and significant correlations have been obtained between Brook score and mean rating. Of the remaining nine correlations, all but one are positive, although none attains significance. Of special interest are the associations for *biological science* and *physical science*. This differentiation is particularly striking among such a highly selected group of subjects, all of them reading psychology.

Table 3. *Relations between* Study of Values *scores and* (a) *Brook score*, (b) *mean ratings for psychology students*

		Correlations with Brook score				Correlations with interest self-ratings			
Study of values category	Brook interest category	n	ρ	t	$P <$ (one-tailed)	n	ρ	t	$P <$ (one-tailed)
Theoretical									
Mean, 34·55;	*I*	80	0·180	1·615	0·1	55	0·220	1·641	0·1
range, 22–47	*SB*	80	0·040	—	—	55	−0·083	—	—
	SN	80	0·286	2·635	0·005	55	0·180	1·616	0·1
	I+SB+SN	80	0·253	2·309	0·025	55	0·159	1·172	—
Aesthetic									
Mean, 33·9;	*A*	80	0·226	2·048	0·025	59	0·536	3·121	0·0005
range, 11–52½	*Lt*	80	0·025	—	—	59	0·382	4·788	0·005
	A+Lt	80	0·215	1·943	0·05	59	0·585	5·444	0·0005
Social									
Mean, 31·46;	*H*	78	−0·014	—	—	57	0·074	—	—
range, 18–46	*Pe*	80	−0·019	—	—	57	0·016	—	—
	H+Pe	80	−0·007	—	—	57	0·055	—	—
Economics									
Mean, 22·51; range, 6½–40	*B*	80	−0·01	—	—	52	0·487	3·941	0·0005
Religious									
Mean, 29·86; range, 4–58	*R*	80	0·182	1·616	0·1	48	0·560	4·586	0·0005
Political									
Mean, 27·92	—	—	—	—	—	—	—	—	—

Table 3 shows the rank correlations between five of the six *Study of Values* measures and (*a*) the relevant Brook scores, (*b*) the mean ratings. Since the Brook interest categories tend to be narrower than those of the Values, we have grouped together the Brook interests which might be analogous, separately or jointly, and have compared them with the appropriate Values scores. Thus, for example, the Values *theoretical* scores fail to correlate significantly with the Brook scores on *intellectual* or *biological science*, but they do correlate significantly with the Brook score on *physical science* and also with these three Brook indices combined.

On *aesthetic* interests, the correlations between the Values score and relevant ratings (Table 3) are higher than those between the Brook scores and ratings (Table 2). It may be seen, however, that significant associations were obtained between the *aesthetic* Brook scores and both the Values scores and the ratings. The Values *social* score fails to correlate with the other criteria, whether *humanitarian* or *people* but *humanitarian* Brook score does correlate significantly with *humanitarian* ratings (cf. Table 2).

The Values *economic* and *religious* scores may be seen to correlate more highly with the relevant Brook ratings than do the Brook scores; in fact, these correlation coefficients are among the highest in the whole table. These latter results are considered in the discussion, below.

Whilst the *political* category in the Brook test is used specifically and narrowly to refer to politics in the sense of current affairs and active membership of political parties, the term is used far more widely in the *Study of Values*, to include power-seeking generally. This particular comparison, therefore, was not considered worth while.

The third test taken by the psychology students was the Neymann & Kohlstedt (1929) test of introversion-extraversion. This consists of fifty questions (of the type 'Do you like to be by yourself a great deal?' 'Do you like to be a leader at a social affair?') and the norms are so arranged that the subject emerges as one of the five following types: 'decided introvert', 'moderate introvert', 'intermediate', 'moderate extravert' or 'decided extravert'. These latter scores were correlated with the Brook D score ('dances, parties and social functions generally') and D ratings, satisfactorily high values of ρ being obtained (see Table 4).

Table 4. *Relations between introversion–extraversion score and (a) Brook* D *score and (b) mean ratings for* D *('dances, parties and social functions generally') for psychology students*

Introversion–extraversion test score		n	ρ	t	$P <$ (one-tailed)
Mean, 0·98;	Brook D score	93	0·265	2·621	0·01
range, −28 to +28	Interest D rating	65	0·430	3·780	0·0005

The following tendencies are of some interest, although they did not reach significance level. (*a*) The 'decided introverts' gained the highest $Un+Mo$ scores. (*b*) The 'decided extraverts' gave few omissions. (*c*) Subjects with (i) a high number of *sinister* (Sin) responses and (ii) a high number of *unclassifiable* (Un) responses tended to gain either extreme introversion or extreme extraversion scores.

B. Naval ratings

Procedure

Seventy-two sailors were given the Brook Reaction test. They were mainly seamen, stokers, electricians and naval airmen. All were junior ratings and none was highly skilled. Fifty of these subjects assessed themselves on the Brook rating sheet (Table 1) but we were unable to obtain a second rating for them. They were usually tested in groups of six, but they had not been together as a group for a long enough period to be able to rate one another. In addition to the Brook test, the sailors took AH 4, a group test of intelligence (Heim, 1955) and a vocabulary test (E of the Self-judging Vocabulary Scale, Heim & Watts, 1961).

Results

Not unexpectedly, sailors differed from undergraduates in having generally less numerous and less strong interests. This was reflected in the Brook scores, sixteen of which yielded lower means for the sailors than for the psychology students. The six exceptional interest categories were B, F, G, K, M and Sl.

The number of significant associations between the sailors' Brook scores and their ratings was considerably lower than that found with the students. This relative lack of agreement is probably due (*a*) to the enforced use of single ratings only and (*b*) to the fact that rating oneself is a technique less suited to sailors than to students. With this group of fifty sailors, only five of the interests yielded significant values of ρ: *C*, *F*, *K*, *Pl* and *Sp*. One of these, however, *F*, had failed to show significant agreement in the student group.

The lower Brook mean scores are due to increased *Un* and *Mo* responses from the sailors. We thought that this might be the effect of a lower degree of intelligence or of literacy on the part of the sailors and, in order to test this hypothesis, we compared their Brook scores with their Intelligence and Vocabulary scores. We argued that if they gained low interest scores owing to such factors as slow speed of writing (resulting in frequent responses of one word only, or even total lack of response) or of poor vocabulary, we should find a significant negative association between number of *Mos* and (*a*) AH 4 score, (*b*) Vocabulary score.

Whilst it is true that both correlations were negative ($\rho = -0{\cdot}101$ and $-0{\cdot}109$, respectively), neither of these approached significance. The only significant association was that between number of multiple-theme responses and AH 4 scores ($\rho = -0{\cdot}255$; $P < 0{\cdot}05$). Thus among the sailors there was a tendency for the *less* intelligent to give more multiple-theme responses. It looks, therefore, as though the Brook test results will not be confused by the degree of intelligence or size of vocabulary of the subject, except perhaps with those of very low intellectual calibre. If the Brook test is to be used with these subjects, separate norms would need to be established using, very likely, a longer time limit. It is relevant to mention the finding of Cantril (1932), quoted by Vernon (1953), 'that persons with high scores on the Study of Values test react *more quickly* to stimulus words connected with their values' (present authors' italics).

C. Criterion groups

Procedure

The criterion groups included the following subjects: a mixed collection of Cambridge undergraduates, male and female; a highly selected group of third and fourth year male students from the Department of Industrial Management; a group of financially-minded graduates, engaged in investment research; a group of London University students reading architecture; a group of women secretaries; and the groups of R.N. engineering students and of sailors described above.

In the case of the Cambridge undergraduates reading a variety of subjects, we took as our criterion membership of university and college societies, clubs and teams. Thus playing for a college or university team, for instance, gave a plus on *sport*; membership of the Student Christian Movement or of the Cambridge Intercollegiate Christian Union gave a plus on *religious*; membership of the Balance Group or treasurership of a Society gave a plus on *business*; and so on. Students could, and in fact always did, gain a plus on more than one category. But a student who belonged to two or more similar societies—for example, two political societies and three college teams—would still merely gain one plus for *political* and *sport*, respectively.

Results

Table 5 (left-hand main column) shows the student membership/non-membership results, for $n > 10$. It may be seen that the difference goes in the predicted direction in eleven cases out of the twelve. Using the Mann–Whitney *U* test, these differences are significant for the interest categories of *business*, *religious*, *biological science*, *physical science* and *sport*.

The other criterion groups confirmed our hypotheses, as may be seen in Table 5 (centre main column). Thus, again using the Mann–Whitney U test, the industrial management and the investment research groups scored significantly more highly on the Brook test in *business* than did the other university subjects; the secretaries scored significantly more highly in *secretarial* than did the female students; the architecture students scored significantly more highly in *aesthetic* than their non-architectural opposite numbers; and the R.N. engineering students scored significantly more highly on *military* than the male students.

The results of comparing the group of sailors with the group of male students are shown in Table 5 (right-hand main column). A two-tailed test was used because we had hoped that the sailors would score more highly on a few of the interests and that the students would score more highly on rather more interest categories. This proved to be the case. The sailors were significantly higher on four of the interests, of which the only surprising one is *secretarial*. Whereas the students came significantly higher on ten of the interests, none of which is surprising, with the possible exception of *travel* which we had not predicted either way.

Despite the higher level of interests of the students, as reflected in Brook scores, the sailors self-rated themselves lower on only ten of the twenty-two categories. For all but one of the ten, the sailors did in fact gain significantly lower Brook scores (the exception, once again, being *secretarial*).

DISCUSSION

The tables show that statistically significant validatory evidence of one kind or another has been gained for twenty of the twenty-two interests (the two exceptions being *entertainment* and *legal*). For some of these categories, such as *aesthetic* and *physical science*, the evidence is fourfold; for others, such as *practical*, it is threefold; for others, such as *clothing and appearance*, it is twofold; and for others, such as *outdoor* and *intellectual* interests, there is at present only one piece of supporting evidence that attains significance. Such variations are due partly to the nature of the interests themselves. It is to be expected that most students have *intellectual* interests: hence, perhaps, the lack of significant differentiation among them on this. However, the difference between the students as a whole and the sailors as a whole is highly significant (U test, $z = 5{\cdot}40$; $P = 0{\cdot}00003$), higher than the difference between the students and the sailors on the Study of Values *theoretical* score (U test, $z = 3{\cdot}26$; $P = 0{\cdot}0007$).

On ratings, however, several of the six Values interests yielded a closer correlation with *Study of Values* score than with the corresponding Brook score. This was true notably in the case of the Values *economic* and *religious* scores and, to a lesser extent, of *aesthetic* score. This may be due largely to the resemblance between the two criteria used: rating oneself is very similar to answering direct, forced-choice questions about oneself. We are glad therefore to have in addition the data of criterion groups, since these are not open to the same objection; and, on criterion groups alone, we have twenty-three significant associations.

These are perhaps most striking with respect to the architecture students, the investment research cum industrial management group, the R.N. engineering

Table 5. *Comparison of Brook test scores for the criterion groups*

In the first two main columns, n is given in brackets. For n, in the third column, see note at foot of table. M = males, F = females

	Membership/non-membership of relevant university societies				University and other sub-groups				Male students and sailors			
Interest category	Group with higher score	Group with lower score	U test		Group with higher score.	Group with lower score	U test		Group with higher score	Group with lower score	U test	
			z	(P 1-tail)			z	(P 1-tail)			z	P (2-tail)
A	Member (21)	Non.m. (29)	1·09	N.S.	Architect students (12 M)	Other students (182 M)	3·43	0·001	Students*	Sailors*	7·31	< 0·0001
B	Member (12)	Non-m. (38)	2·59	0·005	Investment group + Industrial management students (24 M+F)	Other students (263 M+F)	4·15	< 0·0001	Sailors	Students	2·94	0·003
C	—	—	—		—	—	—		Students	Sailors	0·90	N.S.
D	Member (15)	Non-m. (35)	1·05	N.S.	—	—	—		Students	Sailors	1·84	N.S.
E	—	—	—		—	—	—		Students	Sailors	1·06	N.S.
F	—	—	—		—	—	—		Sailors	Students	1·34	N.S.
G	—	—	—		—	—	—		Sailors	Students	1·44	N.S.
H	Non-m. (33)	Member (17)	0·29	N.S.	—	—	—		Students	Sailors	6·09	< 0·0001
I	Member (16)	Non-m. (34)	0·48	N.S.	—	—	—		Students	Sailors	5·40	< 0·0001
K	—	—	—		—	—	—		Sailors	Students	2·91	0·004
Lt	—	—	—		—	—	—		Students	Sailors	5·97	< 0·0001
Lw	—	—	—		—	—	—		Students	Sailors	0·64	N.S.

Table 5 (*cont.*)

In the first two main columns, *n* is given in brackets. For *n*, in the third column, see note at foot of table. M = males, F = females

Interest category	Membership/non-membership of relevant university societies				University and other sub-groups				Male students and sailors			
	Group with higher score	Group with lower score	U test z	U test (*P*1-tail)	Group with higher score	Group with lower score	*U* test z	*U* test (*P* 1-tail)	Group with higher score	Group with lower score	*U* test z	*U* test *P* (2-tail)
M	—	—	—		R.N.E.C. students (56 M)	Other students (138 M)	7·84	< 0·0001	Sailors	Students†	5·06	< 0·0001
O	Member (13)	Non-m. (37)	1·02	N.S.	—	—	—		Students	Sailors	1·42	N.S.
Pe	—	—	—		—	—	—		Students	Sailors	5·52	< 0·0001
Pl	Member (25)	Non-m. (25)	1·16	N.S.	—	—	—		Students	Sailors	4·90	< 0·0001
R	Member (13)	Non-m. (37)	1·71	0·04	—	—	—		Students	Sailors	3·09	0·002
SB	Member (18)	Non-m. (32)	1·86	0·03	—	—	—		Students	Sailors	4·51	< 0·0001
SN	Member (14)	Non-m. (36)	3·24	0·007	—	—	—		Students	Sailors	6·35	< 0·0001
Sl	—	—	—		Secretaries (12 F)	Students (88 F)	3·26	< 0·001	Sailors	Students	4·03	< 0·0001
Sp	Member (23)	Non-m. (27)	2·14	0·02	—	—	—		Students	Sailors	0·56	N.S.
T	Member (14)	Non-m. (36)	0·75	N.S.	—	—	—		Students	Sailors	3·47	< 0·001

* Sample numbers: for sailors, $n = 72$ throughout; for students, $n = 194$ in all cases except that marked †.

† $n = 138$.

students and the secretaries, who gained significantly higher Brook scores, as predicted, on *aesthetic*, *business*, *military* and *secretarial* interests, respectively. We have not, as yet, tested professionals (or amateurs) in the sphere of entertainment, or members (or future members) of the legal profession.

With regard to the comparison between sailors and students, one of the most satisfactory findings was the fact that, despite the relative paucity of the sailors' interests (sometimes recorded by them as 'beer and women' in the 'other specially keen interests' space) the sailors nonetheless gained significantly higher Brook scores on four of the categories. This suggests that the Brook test is not simply reflecting a difference in home background or occupation, with a 'richer' environment automatically producing a higher score.

It may be recalled that one of the most recalcitrant sets of interests proved to be the 'social' ones. The Values *social* score failed to correlate significantly with ratings on *H* (*humanitarian*) or *Pe* (*people*), or on the two combined, and it actually correlated negatively with the corresponding Brook scores. The *Pe* score also correlated negatively with *Pe* mean rating. This suggests that it may be an exceptionally difficult interest to rate or that our description of it is poorly expressed. It is certainly one of the harder interests to score, when interpreting the Brook responses.

H interests also are difficult to score, because it is not clear in all cases 'on whose side' the subject is; for example, in responses such as 'prison' to SELL (CELL) and 'colour' to BAR. Sometimes the subsequent response-word clarifies the situation, but not always. The Brook *H* score did, however, yield a significant association with mean rating; and this score, as well as the *Pe* score, produced highly significant differences between the sailors and the students, the latter proving the more socially inclined.

Wherever it has been possible to investigate individuals further, we have found that a high Brook score is more meaningful than a low Brook score. People who gain an above-average score on any particular Brook interest category tend almost always to be genuinely and deeply interested in the category in question. But a *low* score on a particular interest evidently does not always necessarily imply a complete lack of interest. This is perhaps analogous to results for intelligence tests: a subject who scores highly is at least as intelligent as his score denotes him to be, but a low-scoring subject is not necessarily dimwitted.

The associations between *Un* (*unclassifiable*) plus *Mo* (omitted or single word) responses, multiple themes and *Sin* (*sinister*) responses, on the one hand, and introversion–extraversion scores on the other, are interesting and suggest that further work on these Brook indices might prove fruitful. It is worth mentioning that a student who later had a severe mental breakdown gave as many as four *Sin* responses and obtained one of the highest scores in the group on introversion. It is hoped that future work will include a study of the diagnostic value of the Brook Reaction test for delinquents and neurotics, and an analysis of the interests scores of school children and students from the point of view of educational and vocational guidance.

We should like to express our thanks to Dr M. L. Johnson Abercrombie of the Bartlett School of Architecture, Lt. Cdr. D. J. Mills, R. N., of the Royal Naval Engineering College and Mr A. G. Ellinger of Investment Research for their kind co-operation, and we should like to thank, too, our many subjects. We are indebted to the Medical Research Council for financial support.

APPENDIX

Table 6. *Representative sample of ten out of the eighty stimulus words in the present test*

Serial no.	Stimulus word
8	PLOT
16	TRAIN
24	KEY (QUAY)
32	CARVING (CALVING)
40	RACE
48	PEAK (PIQUE) (PEEK) (PEKE)
56	SERVICE
64	GRAZE (GREYS) (GRAY'S)
72	CHARACTER
80	CLERICAL

REFERENCES

ALCOCK, T. (1963). *The Rorschach in Practice.* London: Tavistock Publications.

ALLPORT, G. W., VERNON, P. E. & LINDZEY, G. (1960). *Study of Values: A Scale for Measuring the Dominant Interests in Personality.* Boston, Mass.: Houghton Mifflin.

BROOK, D. F. & HEIM, A. W. (1960). A preliminary note on the Brook Reaction test. *Brit. J. Psychol.* **51**, 347–56.

CANTRIL, H. (1932). General and specific attitudes. *Psychol. Monogr.* **42**, no. 192.

EYSENCK, H. J. (1964). The biological basis of criminal behaviour. *Advanc. Sci. Lond.*, **21**, no. 92, 338–48.

HARGREAVES, D. H., HEIM, A. W. & WATTS, K. P. (1963). An experiment on the effects of mental set in the Brook Reaction test. *Brit. J. educ. Psychol.* **33**, 236–9.

HEIM, A. W. (1955). *Manual for the Group Test of Intelligence, AH 4.* London: National Foundation for Educational Research.

HEIM, A. W. & WATTS, K. P. (1961). A preliminary study of the Self-judging Vocabulary Scale. *Brit. J. Psychol.* **52**, 175–86.

JUNG, C. G. (1918). *Studies in Word Association.* London: Heinemann.

KENT, G. H. & ROSANOFF, A. J. (1910). A study of association in insanity. *Amer. J. Insanity*, **67**, 37–96.

KENT, G. H. & ROSANOFF, A. J. (1911). A study of association in insanity. *Amer. J. Insanity*, **67**, 317–90.

KUDER, G. F. (1960). *Kuder Preference Record—Personal.* Chicago: Science Research Associates.

MESSICK, S. & ROSS, J. (ed.) (1962). *Measurement in Personality and Cognition.* New York: Wiley.

NEYMANN, C. A. & KOHLSTEDT, K. D. (1929). A new diagnostic test for introversion–extraversion. *J. abnorm. soc. Psychol.* **23**, 482–7.

SIEGEL, S. (1956). *Nonparametric Statistics.* New York: McGraw-Hill.

STRONG, E. K., JR. (1959*a*). *Strong Vocational Interest Blank for Men.* Palo Alto: Consulting Psychologists Press.

STRONG, E. K., JR. (1959*b*). *Strong Vocational Interest Blank for Women.* Palo Alto: Consulting Psychologists Press.

VERNON, P. E. (1953). *Personality Tests and Assessments.* London: Methuen.

(*Manuscript received* 30 *July* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 187–192
Printed in Great Britain

PREDICTION OF GRID SCORES

By J. M. M. MAIR
Middlesex Hospital Medical School, London, W. 1

The grid method, developed by Kelly (1955), is supposed to provide meaningful measurements of the relations between the constructs that any individual uses. In this study an attempt has been made to assess the validity of this claim within the area of public constructs (or dictionary meaning) rather than within the more complex domain of personal or private constructs. The adequacy of the grid method in objectifying meaningful construct relations, changes and stability is examined. Thirteen medical students acted as subjects. From a total of 61 predictions, all but 4 were fully confirmed. The results therefore tend to support claims for the grid's possible usefulness in relation to each of the three functions examined.

The grid method was developed by Kelly (1955) for measuring relations between the constructs any individual uses. In recent years, modifications of this method have been used in a number of experimental investigations (Bannister, 1960, 1962*a*, 1963, 1965*a*; Mair, 1964*a*) and clinical investigations (Bannister, 1965*b*; Fransella & Adams, 1966; Crisp, 1964). Direct evidence concerning different aspects of the reliability and validity of grid measures is scarce. This may be due partly to the fact that the grid is a variable technique (in which different elements and constructs can be inserted for different purposes) and not a static test; because of this, it is impossible to make a final assessment of reliability or validity.

Kelly's theory has been described elsewhere (Kelly, 1955; Bannister, 1962*b*) and the close relation between the theory and the grid method has been discussed by these writers and by Mair (1964*b*). Kelly implies that the same word label may be applied by different people to rather different things, and the relation between similarly labelled constructs for two people may be different. This makes it difficult to assess whether grid scores from the interrelations between any set of personal constructs accurately reflect the peculiarities of the individuals involved, or merely indicate some unsatisfactory aspect of the grid method itself. Furthermore, different external circumstances may be required to make different people modify relations between their personal constructs since they may interpret any particular situation in different ways. It would be difficult therefore to produce similar changes in construct relations for different people, or to assess the meaningfulness of the changes, or lack of them, in grid scores which might emerge from such an attempt. Before embarking on such a project some simpler means of estimating the capacity of the grid method to reflect meaningful relations, changes and stability, is needed.

People function psychologically both within their own personal construct system and also within a public system of constructs, i.e. of dictionary definitions of constructs. This public system is more stable and more open to inspection than the personal one. In this study, parts of the public system have been used to allow assessment of the validity of grid scores as measures of meaningful construct relations, change and stability. The aims of the investigation were, therefore: (1) To assess whether the grid method allows reasonably accurate objectification of relations between constructs. (2) To find if known changes in construct relations are reflected in appropriate grid-score changes. (3) To find if construct relations which are unaltered are reflected in stable grid scores.

METHOD

Rationale

In the personal use of constructs no *right* or *wrong* relations between constructs exist. Relations can only be considered as more or less useful for the individual in his predictive efforts. Within the public system (of dictionary definitions) degrees of accuracy and inaccuracy in the use of constructs can be estimated. Constructs with public definitions may still, of course, be used with private meanings if a person is not familiar with the correct public usage.

If two adjectives (which can be used to describe people) with similar public meanings are chosen, and used as constructs in a grid, then, for a group of people, high positive relations between them should emerge in terms of grid score matchings, provided that the subjects know the public meaning of each adjective. If any subject does not initially know the meaning of one of the adjectives, he may use the adjective in a personalized sense (i.e. guess the meaning). The mean grid relation between the two adjectives for a group of such people should be near the chance level. If each subject thereafter discovers the correct public meaning of the unknown word, repetition of grid testing should reveal an overall increase in relation between the two constructs. Also on retesting, relations between constructs whose meanings were originally known should remain stable.

Each of these expectations should be supported provided that the grid method reflects the appropriate characteristics of the construct systems involved. Support for these predictions would suggest the possible value of grid scores in the area of personal construct usage, where the *correct* scores could not be assessed.

The grid

A modification of the original grid method introduced by Bannister (1962*a*) has been used. Twenty passport-type photographs of people unknown to the subjects were used as elements. All the constructs were supplied to the subjects by the experimenter. For each construct (descriptive adjective) the subject was asked to select the ten people (photographs) who in his opinion most certainly showed that characteristic.

Twelve constructs were used. There were five pairs of synonyms or near synonyms in addition to the constructs CONFIDENT and LIKE ME. (These two constructs were included primarily to provide a cross-check on findings from another study, but they have been included in some of the present calculations). One of the pairs included the constructs ANXIOUS and EASILY UPSET. These were thought to be similar in meaning, and easy enough to be understood by all subjects. In each of the other four pairs one of the words was thought to be relatively easy and familiar, and the other more difficult, though of similar meaning. The 'difficult' words in each pair were selected from the Senior Mill Hill Vocabulary Test (Raven, 1948). The 'easy' items were the equivalents provided in the synonyms section of the Mill Hill test. These four pairs are referred to throughout as the 'easy/difficult' pairs. With the 'easy' word first, they were DILIGENT/SEDULOUS, POMPOUS/BOMBASTIC, MENACING/MINATORY and PETULANT/QUERULOUS.

The grid was administered twice using the same elements and the same constructs. On each occasion the order of the constructs was as follows: CONFIDENT : DILIGENT : POMPOUS : MINATORY : SEDULOUS : BOMBASTIC : PETULANT : ANXIOUS : MENACING : QUERULOUS : EASILY UPSET : LIKE ME. The scoring method and its rationale are given fully by Bannister (1960). In a grid with twenty elements there are eleven possible scores going by steps of two from 0 to 20.

Mill Hill Vocabulary Test

The Form 2 Senior, Definitions section only, was used.

Procedure

Testing was done in two sessions. In the first, each subject completed the Definitions section of the Mill Hill Test. This allowed assessment of whether or not each subject had some idea of the meaning of the difficult words in four of the pairs. The subject was then introduced to the grid as a test of his ability accurately to describe the people shown in a set of photographs in terms of various adjectives. The grid was then administered in the usual way. The subject was told to guess or make up a meaning for the adjectives he did not know. He was told that the adjectives could be used to describe certain characteristics people have.

After doing the grid, each subject was asked to explain the meaning he had attributed to the adjectives he did not know. When assessing whether any subject knew the meaning of any of the difficult items, both his Mill Hill and grid usages of the terms were considered.

At the end of the first session, each subject was given a list of all the words in the easy/difficult pairs, and asked to check their meanings before returning for the second session. At the beginning of the second, each subject was asked to give the correct meaning of each of the words. The grid was then repeated. The average period between the first and second session was 11 days.

Subjects

Thirteen medical student volunteers took part (five female; eight male).

Predictions

Specific predictions related to the three main aims were tested.

(1) Matching scores and meaning

(*a*) Within the easy/difficult pairs on the first session, the grid relations between pairs where the subject knows the meaning of the 'difficult' word will be significantly higher than the grid scores between pairs where the 'difficult' word is not known.

(*b*) There will be a positive relation between ANXIOUS and EASILY UPSET, giving an average matching score of 14+ on the first occasion.

(*c*) Within the easy/difficult pairs, the average matching score, between pairs where the meaning of the 'difficult' word is not known, should be at the chance level (= 10) on the first occasion.

(*d*) Within the easy/difficult pairs, the average matching score between pairs should increase ($\geqslant 14$) on the second occasion.

(2) Changes in construct relations.

(*a*) Within the easy/difficult pairs, between the first and second sessions, the grid scores for each subject, and for each of the four pairs should increase (or remain at the same level if originally high, i.e. $\geqslant 14$).

(*b*) Any variation in matching scores between constructs, from the first to the second testing sessions (other than in the easy/difficult pairs), should be distributed on a chance basis between increases, decreases, and scores at the same level.

(*c*) Any variation in matching scores between constructs, from the first to the second sessions (other than for the easy/difficult pairs), should, overall, be smaller than matching score changes in the easy/difficult pairs.

(3) Stability of Construct Relations

Within the easy/difficult pairs, the change in matching scores between pairs where both words were known on the first occasion, and between pairs where one or both were unknown should be statistically significant. The known pairs should be more stable.

RESULTS

(1) *Matching scores and meaning*

(*a*) In Table 1 the mean of the grid matching scores of the word pairs where each subject had some idea of the meaning of the more 'difficult' word is compared with the mean of the grid matching scores between the pairs where he or she apparently did not know the correct meanings of the 'difficult' item. The assessment of meaning was done prior to scoring the grids, and quite independently of grid scores. Using Wilcoxon's tests for matched pairs (Siegal, 1956) the difference in grid scores from pairs where the meaning was known and pairs where the meaning was not known was highly significant ($T = 0$, $n = 9$; $P < 0{\cdot}005$, single-tail; subjects 2 and 10 were excluded from the calculation since they provided no comparative data—one knew all the meanings, the other knew none).

(*b*) The mean matching score between ANXIOUS and EASILY UPSET in the first session was 14·8. Eleven of the thirteen subjects gained scores of 14 or above.

(*c*) Within the easy/difficult pairs, the average matching score on the first occasion between pairs where the meaning of the difficult word was not known was 10·3.

Table 1. *Mean grid scores between pairs where the meaning was known, and pairs where the meaning was not known*

Subject	Meaning known	Meaning unknown
1	14	8
2	13	—
3	14	14
4	13	6
5	13	10
6	14	9
7	12	12
8	13	9
9	14	13
10	—	12
11	12	11
12	18	10
13	16	10

(*d*) Within the easy/difficult pairs, the average matching score on the second occasion was 15·4.

(2) *Changes in construct relations*

(*a*) For the easy/difficult pairs, forty-eight matching scores increased in size or remained high and positive ($\geqslant 14$) from the first to the second grid sessions, four decreased or remained low. This distribution is obviously significantly different from chance expectancy.

(*b*) Table 2 shows a comparison of changes in matching scores from pairs where a change in the direction of increased scores was expected, and from two sets of four matchings where no general trend was anticipated.

Control (1) involves matchings, from first and second sessions, of LIKE ME with CONFIDENT, ANXIOUS, and EASILY UPSET; also ANXIOUS with EASILY UPSET. Control (2) involves matchings, from first and second occasions, of LIKE ME with POMPOUS, PETULANT, MENACING and DILIGENT. The experimental comparison involves matching scores from the four easy/difficult pairs.

Table 2. *Changes in matching scores for the easy/difficult pairs and two sets of control data*

	Increase in matching scores	Same matching scores	Decrease in matching scores
Experimental comparison	41	8	3
Control (1)	15	20	17
Control (2)	20	14	18

Both control comparisons show a chance distribution of changes (χ^2 test). The experimental comparison shows a highly significant change in the predicted direction ($\chi^2 = 43{\cdot}43$, D.F. $= 2$; $P < 0{\cdot}001$).

(*c*) The mean changes in matching scores for the groups were as follows: experimental, 4·2; control (1), 1·8; control (2), 2·2. The overall change shown by the experimental pairs is significantly greater (Wilcoxon Test) than that for the other two groups (for E *v.* C(1), $T = 3{\cdot}5$, $n = 13$, $P < 0{\cdot}005$; for E. *v.* C(2), $T = 4{\cdot}5$, $n = 12$, $P < 0{\cdot}005$; for C(1) *v.* C(2), $T = 29{\cdot}5$, $n = 13$, N.S.).

The mean differences in fact minimize the actual changes in grid scores which did emerge because a considerable number of subjects gained high positive matching scores in the first session between POMPOUS/BOMBASTIC and PETULANT/QUERULOUS. Excluding the former, the mean change in matching scores for the experimental pairs was 5·0. If both comparisons are excluded, the mean change was 6·0.

(3) *Stability of construct relations*

Table 3 shows the mean changes in each subject's matching score from first to second occasions for easy/difficult pairs where the difficult word was known originally, and for pairs where the meaning was not known. The difference between these distributions of differences was highly significant (Wilcoxon's $T = 0$, $n = 9$, $P < 0{\cdot}005$). As predicted, the changes in grid scores for pairs where the meaning was originally known was significantly less than for pairs where the meaning was not known. The mean change for all pairs where the meaning was known was 2·09; when the meaning was originally unknown, the overall mean change was 5·51.

Table 3. *Mean changes in each subject's matching score for pairs where the meaning was originally known and pairs where the meaning was not known*

Subject	Mean changes where meaning known	Mean changes where meaning not known
1	2·0	4·0
2	4·5	—
3	2·0	2·0
4	2·0	7·0
5	1·0	4·0
6	2·0	7·3
7	4·0	4·7
8	2·0	7·0
9	2·0	3·3
10	—	3·0
11	2·0	7·3
12	2·0	6·0
13	2·0	8·0

DISCUSSION

The main aim of the present study was to set up an experimental situation in which the adequacy of certain aspects of the grid method could be examined. The questions asked were, 'Does the grid method allow accurate and objective assessment of (1) meaningful relations between constructs, (2) meaningful and expected changes in construct relations, and (3) stability in construct relations where no changes were intended?' From a total of sixty-one predictions, all but four were fully confirmed. The results therefore tend to support usefulness of the grid in relation to each of the three functions examined.

The study was not intended to be definitive. Its results indicate that under controlled circumstances grid scores tend to reflect *predicted* relations, changes and stability. Some predictions were not fully supported, and in further experimentation closer attention may need to be paid to the actual processes involved when a subject selects elements on any construct dimension. This kind of detailed exploration of individual subjects is likely to contribute more towards understanding the ways in which construct systems function than large-scale studies concerned with the overall reliability or changeability of grid scores.

REFERENCES

BANNISTER, D. (1960). Conceptual structure in thought disordered schizophrenics. *J. ment. Sci.* **106**, 1230–49.

BANNISTER, D. (1962*a*). The nature and measurement of schizophrenic thought disorder. *J. ment. Sci.* **108**, 825–42.

BANNISTER, D. (1962*b*). Personal construct theory: a summary and experimental paradigm. *Acta psychol., Amsterdam* **20**, 104–20.

BANNISTER, D. (1963). The genesis of schizophrenic thought disorder: a serial invalidation hypothesis. *Brit. J. Psychiat.* **109**, 680–6.

BANNISTER, D. (1965*a*). The genesis of schizophrenic thought disorder: re-test of the serial invalidation hypothesis. *Brit. J. Psychiat.* **111**, 377–82.

BANNISTER, D. (1965*b*). The rationale and clinical relevance of repertory grid technique. *Brit. J. Psychiat.* **111**, 977–82.

CRISP, A. H. (1964). An attempt to measure an aspect of transference. *Brit. J. med. Psychol.* **37**, 17–30.

FRANSELLA, F. & ADAMS, B. (1966). An illustration of the use of repertory grid technique in a clinical setting. *Brit. J. soc. clin. Psychol.* **5** (in the Press).

KELLY, G. A. (1955). *The Psychology of Personal Constructs.* vol. I. New York: Norton.

MAIR, J. M. M. (1964*a*). Paired associate learning in neurotics: an investigation of individual differences and the effects of repeated failure. Unpubl. Ph.D. thesis, University of London Library.

MAIR, J. M. M. (1964*b*). The derivation, reliability, and validity of grid measures: some problems and suggestions. *Bull. Brit. psychol. Soc.* **17**, no. 55, 7A.

RAVEN, J. C. (1948). *Guide to Mill Hill and Matrices.* London: H. K. Lewis.

SIEGAL, S. (1956). *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill.

(*Manuscript received* 10 *November* 1965)

Brit. J. Psychol (1966), 57, 1 *and* 2, *pp.* 193–199
Printed in Great Britain

GEOMETRIC ILLUSIONS AND ENVIRONMENT: A STUDY IN GHANA

By GUSTAV JAHODA
University of Strathclyde, Glasgow

Segall, Campbell & Herskovits (1963, 1966) put forward the view that the Müller–Lyer illusion is a function of rectangularity in the environment, and that the horizontal–vertical one depends on openness of terrain. Existing contrasts in these two variables in Ghana were used in an attempt to investigate the hypotheses. A total of 213 illiterate subjects were tested, and 41 subjects in Britain. A significant overall difference in the expected direction between Ghanaian and British subjects was obtained with the Müller–Lyer, but no differences corresponding to environmental variations emerged on either of the illusions within the subgroups of Ghanaian subjects. The reasons for this partial failure to replicate are examined in the light of other studies, and some theoretical implications discussed.

Half a century after Rivers (1901, 1905), interest in the cross-cultural study of visual illusions has revived as a result of the initiative by Segall *et al.* (1963, 1966). By means of carefully devised stimulus material they obtained responses to illusions from samples of Europeans, Africans and some other cultures. Their findings were, in brief, that European samples were more susceptible to the Müller–Lyer (M–L) and Sander Parallelogram illusions than non-European samples; on two forms of the horizontal–vertical (H–V) illusions, by contrast, most non-European samples proved more susceptible. Their explanation of these results is in terms of 'learned habits of inference' based on variations in physical and cultural environments. More specifically, the following processes were postulated by Segall *et al.* (1966).

(1) 'For figures constructed of lines meeting in non-rectangular junctions, there will be a learned tendency among persons dwelling in carpentered environments to rectangularize these junctions, to perceive the figures in perspective, to interpret them as two-dimensional representations of three-dimensional objects'.

(2) 'The horizontal–vertical illusion results from a tendency to counteract the foreshortening of lines extended into space away from a viewer, so that the vertical in the drawing which is the stimulus for the illusion is interpreted as representing a longer line. Since the tendency has more ecological validity for peoples living mostly outdoors in open, spacious environments, it is predicted that such peoples will be *more* susceptible than Western peoples in urban environments. On the other hand, some non-Western persons should be *less* susceptible to the illusions, e.g. rain forest or canyon dwellers'.

The general trend of their findings is of course broadly consistent with these interpretations, but the fit is not very close, especially as regards the H–V illusions. The authors were well aware of this, and in connexion with the latter Campbell (1964, p. 316) has written: 'If this ecology is a factor, as I do believe it is, it is obviously only one of many factors.' In their extensive exploratory survey, Segall *et al.* (1966) perforce left many of the factors uncontrolled, one of the major ones being literacy. Its potential importance may be indicated with reference to their first hypothesis, concerning the M–L illusion, which contains two elements: the 'carpentered environment', and the interpretation of two-dimensional drawings. Although closely associated in western cultures, these are in fact distinct, and ought to be treated as such. There is thus a need for some cross-validations of these hypotheses, and the present study is an attempt to make use of the geographical as well as cultural variations in Ghana for this purpose.

METHOD

Subjects

In Northern Ghana, the Lobi and Dagomba are tribes closely similar in cultural characteristics and physical environment. They live in open parkland in rolling country. The huts of both are round, the main difference being that those of the Dagomba tend to be considerably larger; both kinds are usually devoid of rectangular furnishings.

The Ashanti in central Ghana are geographically and culturally quite distinct. They live in a dense forest belt, and remote villages enclosed in high tropical rain forest were specially singled out. Their houses are rectangular, though the edges are not always sharply defined; they contain a good deal of 'carpentered' furniture, tables and chairs being well-nigh universal.

From the point of view of design, it would have been desirable to test both literates and illiterates; but in northern Ghana the bulk of the adult population is illiterate, and the same is true of Ashanti outside the urban areas. Hence testing was confined to illiterates. Subjects were obtained by getting the chiefs or elders to assemble the people in the homestead or village, and from among these a roughly equal proportion of men and women within the approximate age-range 16–50 yr were selected.

In addition to the Ghanaian sample, forty-one subjects were also given the illusions in Britain. They were undergraduates not taking Psychology, and university administrative staff.

Materials

Since the aim of the work was to test the hypotheses put forward by Segall *et al.* (1966), it would have been best if their own materials could have been used. Unfortunately circumstances precluded any possibility of these being obtained in time. The conventional forms of the Müller–Lyer and horizontal–vertical (⊥) illusions, incorporating movable slides, were therefore employed. These were constructed in white plastic with black lines, according to the dimensions specified by Fraisse (1963). In view of the objections raised by Segall *et al.* (1966) to the work of Bonte (1962) it should be mentioned that, apart from the fact that there is no grain as in the wooden device, the illusions were wholly enclosed in a plastic frame, so that there was little apparent discontinuity between the stationary and moving parts. Nonetheless, careful precautions were taken to ensure that the comparisons to be made were correctly understood. The procedures were tried out on illiterate subjects in the southern coastal region, with the assistance of the interpreter who subsequently took part in the actual experiment.

The pilot exercise clearly showed the need to keep the number of trials demanded of any one subject to an absolute minimum. The tolerance of illiterates for a strange task, whose purpose is obscure, is limited; unless one has some special claim on their co-operation, fatigue and boredom set in fairly soon and the performance then becomes perfunctory. In the course of the pilot study a slight mishap also occurred when one end of the black line of the H–V illusion was damaged; hence it was only possible to start from the shorter position. This means that the absolute amount of the illusion effect could not be validly determined, but the group comparisons for testing the hypotheses are of course no affected. The number of trials for each subject was as follows: six for the Müller–Lyer illusion (three in and three out) and five for the horizontal–vertical illusion (out only). Sequence of presentation was randomized throughout.

Procedure

Subjects were told at the outset that we wanted to find out 'how well they could see'. The subject was seated and given the apparatus to hold flat on his lap, at a viewing distance of some 14–18 in. He was allowed to familiarize himself with it by moving the slide in and out. The interpreter then told him that he should make 'the two things like twigs the same', simultaneously demonstrating: for the M–L illusion, two straight twigs were held first end to end, and then in parallel to show their identical length. For the H–V illusion, one twig was first held up horizontally or vertically and turned 90°. The writer, sitting next to the subject, then ran his index from one end of the standard to the other; in the case of the M–L illusion, he also indicated (either in the 'in' or 'out' position, as appropriate) the line from the end of the arrow on the left to the point where the line joined the Y, being careful to stop just short of the junction. Next, subjects were asked to do what they had just been shown; if they went to the end of the total Y part of the M–L illusion, or confined themselves to half the base line with the H–V one, their results were discarded. In order to avoid causing offence, such subjects were actually allowed to make some responses, and rewarded like all others with a small packet of biscuits. This was the only stage at which subjects were eliminated; the loss amounted to about one in twenty, rates being similar in the different tribes.

The use of the method of average error has the drawback that one obtains responses influenced in part by movement, and not purely perceptual ones as those of Segall *et al.* (1963, 1966); furthermore, their stimulus material was presented in a different spatial orientation, namely vertically at eye level. Accordingly, direct comparisons of the extent of illusion effects would be unwarranted. The aim was thus confined to undertaking group comparisons designed to test the two hypotheses below.

Hypotheses

I. Lobi and Dagomba, living in open country, will be *more* susceptible to the H–V illusion than Ashanti, whose surroundings consist of dense tropical forest.

II. Ashanti, living in relatively more 'carpentered' surroundings, will be *more* susceptible to the M–L illusion.

RESULTS

In analysing the data, the median was used as measure of central tendency so as not to give undue weight to occasional extreme responses. The medians were transformed into percentage discrepancies from the standard, and the arithmetic means of these discrepancies computed. When Bartlett's test was applied to the H–V illusion data, a significant heterogeneity of variance ($P < 0{\cdot}02$) was found; hence all data were analysed by the median test, using a two-tailed χ^2 test of significance. No significant sex differences emerged for either of the illusions with any of the tribes, and responses from men and women were therefore pooled.

Horizontal–vertical illusion

On examining the responses to this illusion it became immediately evident that Lobi and Dagomba could not, as originally intended, be treated as a single population. Accordingly they are shown separately in Table 1. The performance of Europeans is near the midpoint of the total range, differing significantly from that of both Lobi and Ashanti ($P < 0{\cdot}01$ and $P < 0{\cdot}02$ respectively). The fact that Europeans do not fall outside the range of the African samples is consistent with the global aspects of the findings by Segall *et al.* (1966), but the pattern within the African samples does not fit their ecological interpretation. The value for Ashanti (tropical forest) falls between those of the two others (both open parkland). At one extreme the Dagomba

Table 1. *Mean percentage discrepancies in the horizontal–vertical illusion*

Sample	*n*	Mean percentage	S.D.
Lobi	34	1·65	13·95
Europeans	41	11·16	8·63
Ashanti	127	16·22	9·74
Dagomba	52	22·71	10·34

Median test: $\chi^2 = 39{\cdot}59$; D.F. $= 3$; $P < 0{\cdot}001$.

gave the largest illusion responses; at the other, the Lobi had a mean percentage discrepancy close to zero; over half of them in fact either gave no illusion response or even a negative one, whereby the vertical segment is made longer than the horizontal one. These results thus provide no support for hypothesis I.

Müller–Lyer illusion

The picture presented by Table 2 is totally different. All the mean percentage discrepancies are closely similar, so that hypothesis II also fails to find support. On the other hand, it appears from the comparison between the combined African versus the European group that the latter exhibits a significantly greater illusion effect, which is in accordance with the postulates of Segall *et al.* (1966).

Table 2. *Mean percentage discrepancies in the Müller–Lyer illusion*

Tribe	*n*	Mean percentage	S.D.
Lobi	34	15·18	8·46
Dagomba	52	17·44	8·15
Ashanti	127	17·61	7·93
Europeans	41	23·22	6·94

Median test: between the three tribes—$\chi^2 = 3{\cdot}91$; D.F. 2; $0{\cdot}20 > P > 0{\cdot}10$; combined tribes versus Europeans—$\chi^2 = 7{\cdot}48$; D.F. 1; $P < 0{\cdot}01$.

DISCUSSION

Since the findings were not in conformity with the hypotheses put forward by Segall *et al.* (1966), the question of how the negative outcome is to be interpreted arises. The first and most obvious possibility would be to attribute it merely to the differences in the methods employed. Some of these have already been mentioned, and it should be added that Segall *et al.* (1966) had an elaborate scheme for eliminating non-scale and inconsistent sets of responses; this was of course essential, since the bulk of their testing was done indirectly. Owing to the limited number of trials per individual subject in the present study, elimination of subjects was confined to the initial stage of the testing session, supervised throughout by the same experimenter; thereby external control of the uniformity of administration in all three tribal areas was ensured.

However, in the case of the M–L illusion such external control would not constitute a fully reliable safeguard against one gross source of error which Segall *et al.* (1966) have regarded as invalidating the work of Bonte (1962). They suggest that Bonte's subjects were led both by the nature of her apparatus (similar to the one used in the present study) and by the instructions to misinterpret their task as a matching of the *total figure* on the left ↔ with the *total figure* –< on the right; this would of course have resulted in a crass over-estimation of the illusion effect. The precautions taken in the present study to obviate this have been indicated previously; but since the illusion effects in the present study were substantially higher than most of those found by Segall *et al.* (1966) in their African samples, it might be objected that the same spurious factor may have been operative; it is therefore necessary to consider this argument.

For this, purpose let us suppose that this factor *was* operative, so that the two total figures were mistakenly matched. If one assumed, for the sake of simplicity, that this matching was on the average free of error, a spurious illusion effect amounting to 25 % would result in each case. Given this base line, one can go on to try out various ways in which the results of the present study, namely a mean value of some 17% for the M–L illusion, could have come about. Clearly, not every subject could have matched total figures, or the mean would be in the region of 25 %; if half the subjects had matched total figures, and the other half had matched in the correct manner but with a nil illusion effect, the net resultant would be an overall mean illusion effect of some 12 %, which is too low; a more realistic assumption would be that the other half showed a moderate genuine illusion effect of some 9 %; this would yield an overall mean of about 17 %, close to the value actually obtained. If something like this had in fact happened one would predict *bimodality* of the distribution of responses, with two major clusters around 25 % and 9 % respectively. The actual distribution is shown in Table 3 to be grouped around the mean; there is no indication whatsoever of bimodality, and it is therefore highly unlikely that the results of the present study can be accounted for by assuming mistaken matching.

Table 3. *Distribution of percentage discrepancies for the Müller–Lyer illusion for the three tribes combined*

			Mean value				
% illusion	–7	8–12	13–17	18–22	23–27	28–	Total
n	21	35	50	51	37	19	213

Apart from this, the difference in magnitude in the M–L illusion effects may well be in part a function of the different methods employed; Segall *et al.* (1966) themselves have admitted that their method tends to weaken the effect somewhat. In any case, it has already been stated that the concern of the present study was to test predicted group differences, not to establish absolute amounts of illusion effects.

In the case of the H–V illusion, however, it must be pointed out that the spatial orientation of the apparatus may have had some bearing on the outcome. As the illusion was presented horizontally, the 'vertical' line was in fact only symbolically vertical, whereas with Segall *et al.* (1966) it was actually so. There seems to be no definite evidence concerning such a position effect, which calls for further investigation.

Could any other elements have contributed to the findings? One possibility that suggests itself is that the overall range of the variables, namely rectangularity and openness of terrain, may have been too restricted. This no doubt may have reduced the effect, but one would still have anticipated some trend in the expected direction.

At this point it is perhaps desirable to view the problem in a wider perspective, beginning with the M–L illusion. In the introduction it was noted that Segall *et al.* (1966) failed to distinguish between a 'carpentered environment' and 'the interpretation of two dimensional drawings', both of which form part of their hypothesis. There is considerable evidence for the inability of illiterates to interpret two-dimensional representations of three-dimensional objects. Biesheuvel (1952) has reported on this; Hudson (1960) carried out a detailed study, which is referred to by Segall *et al.* (1966), though they apparently remained unaware of its full implications. They certainly do not seem to have obtained any details of the cultural and especially the educational background of their individual subjects, relying instead on global information about culture and ecology.

On the basis of the distinction, one might venture to propose that illusion effects of the M–L type may be a combined function of two distinct variables: (1) the degree of rectangularity in the environment, which is the primary one; (2) the extent of the ability to interpret two-dimensional representations. It may well be that in the absence of (2), or with a deficiency in it, relatively moderate variations in (1) produce such limited changes in habits of perceptual inference that the difference cannot be detected by the relatively crude methods employed in the present study. It is even possible that, as far as test responses are concerned, radical differences in (1) may not manifest themselves at all if there is a marked deficiency in (2). At any rate, there is some evidence for this from a study by Mundy-Castle & Nelson (1962) of a backward community of white forest workers in South Africa. The environment of these forest workers is far more 'rectangular' than that of the Bushmen; yet they resemble the Bushmen in (2), since none of them was able to give consistent three-dimensional responses on the Hudson pictorial perception test. Their responses on the M–L illusion did not differ significantly from those of two samples of Bushmen, but were very significantly smaller than those of ordinary white adults. In view of Gregory's (1963) theory, it is also of interest to note that the forest workers showed a significant tendency to underconstancy.

In his article Gregory refers to the findings that for people living in houses without corners the geometrical illusions are reduced; however, this does not apply to the H–V illusion. A variety of at least partial explanations of this type of illusions have been suggested by various workers such as Künnapas (1955, 1957, 1958). None of the theories familiar to the present writer help to throw any light on the totally unexpected contrast found in the present study, with one possible exception: Kohler & Pissarek (1960) put forward, it seems independently, a hypothesis identical with that of Segall *et al.* (1966). They arrived at this in the course of experiments in which subjects wore glasses producing artificial astigmatism. This suggests what is admittedly not more than a speculation: since there is rather rigid endogamy within the tribes studied, there might be a difference in genetically determined astigmatism which could outweigh ecological variations. This would be consistent with the exceptionally high standard deviation of H–V illusion responses among the Lobi, although any other factor tending to dichotomize the population would of course have a similar effect. Moreover, it has been suggested to the writer that permanent astigmatism would probably become compensated, thus not leading to any distortion. Again, this problem appears to be open to specific test.

Whatever the causes of the contrasts encountered, it is worth pointing out that the present findings concerning the H–V illusion are not the only ones inconsistent with a simple ecological interpretation; this is evident from the comment by Campbell (1964), quoted earlier, but a more directly comparable instance may be cited; two successive expeditions to the Kalahari by the same team used the tests of Segall *et al.* (1966) with two groups of Bushmen; on one version of the H–V illusion results were substantially the same, with the other version they differed very significantly ($P < 0{\cdot}01$); here again both tribe and ecology were constant (based on data contained in Morgan, 1959; Reuning, 1959; Mundy-Castle & Nelson, 1962).

One final aspect of the present findings may be briefly noted; it relates to the possible effect of later experience. All subjects in Ghana were asked if they had travelled elsewhere in the country; it emerged that nearly all the men had occasionally visited urban areas; some had worked as seasonal labourers for varying periods, others were employed for spells in the mines. By contrast, nearly all the women had spent their lives in the village or settlement, rarely going beyond a radius of a few miles. In spite of this, there were no significant sex-differences in the responses to either of the illusions. Thus such environmental factors as were operative must have been confined to childhood or adolescence.

In conclusion, the present findings highlight the complexity of the problem, which is apt to be obscured by the tendency of investigators to view the issue in terms of Europeans versus non-Europeans. More specific hypotheses are needed, that can stand the test of accounting for differences within, as well as between, these categories.

Thanks are due to Prof. J. H. Nketia, Department of African Studies, University of Ghana, who generously provided the necessary facilities for the study; I am also indebted to Mr R. L. Gregory, Cambridge University, for valuable criticisms and suggestions.

REFERENCES

BIESHEUVEL, S. (1952). The study of African ability. Part II. A survey of some research problems. *African Studies*, **11**, 105–17.

BONTE, M. (1962). The reaction of two African societies to the Müller–Lyer illusion. *J. soc. Psychol.* **58**, 265–8.

CAMPBELL, D. T. (1964). Distinguishing differences of perception from failures of communication in cross-cultural studies. In *Cross-Cultural Understanding: Epistemology in Anthropology*, ed. F. C. S. Northrop and H. H. Livingston. New York: Harper and Row.

FRAISSE, P. (1963). *Manuel pratique de psychologie expérimentale.* Paris: Presses Universitaires.

GREGORY, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature, Lond.*, **199**, 678–80.

HUDSON, W. (1960). Pictorial depth perception in sub-cultural groups in Africa. *J. soc. Psychol.* **52**, 183–208.

KOHLER, I. & PISSAREK, T. (1960). Brillenversuche zur Vertikalentäuschung. *Psychol. Beitr.* **5**, 117–40.

KUNNAPAS, T. M. (1955). Influence of frame size on apparent length of a line. *J. exp. Psychol.* **50**, 168–70.

KUNNAPAS, T. M. (1957). The vertical–horizontal illusion and the visual field. *J. exp. Psychol.* **53**, 405–7.

KUNNAPAS, T. M. (1958). Influence of head inclination on the vertical–horizontal illusion. *J. Psychol.* **46**, 179–85.

MORGAN, P. (1959). A study in perceptual differences among cultural groups in southern Africa, using tests of geometric illusions. *J. nat. Inst. personnel Res.* **8**, 39–43.

MUNDY-CASTLE, A. C. & NELSON, G. K. (1962). A neuropsychological study of the Knysna forest workers. *Psychologia Africana* **9**, 240–72.

REUNING, H. (1959). Psychologische Versuche mit Buschleuten der Kalahari. *Die Umschau in Wissenschaft und Technik*, no. 17, 520–3.

RIVERS, W. H. R. (1901). Vision. In *Reports of the Cambridge Anthropological Expedition to the Torres Straits*, vol. II, part 1, ed. A. C. Haddon. Cambridge University Press.

RIVERS, W. H. R. (1905). Observations on the senses of the Todas. *Brit. J. Psychol.* **1**, 321–96.

SEGALL, M. H., CAMPBELL, D. T. & HERSKOVITS, M. J. (1963). Cultural differences in the perception of geometric illusions. *Science* **139**, 769–71.

SEGALL, M. H., CAMPBELL, D. T. & HERSKOVITS, M. J. (1966). *The Influence of Culture on Visual Perception.* Indianapolis: Bobbs-Merrill (in the press).

(*Manuscript received* 5 *July*, 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 201–202
Printed in Great Britain

INDUCTION OF STEREOSCOPIC DEPTH EFFECTS

By NICHOLAS PASTORE and MARLENE TERWILLIGER
Queens College, City University, New York 11367

When a pair of continuous horizontal lines of equal length is presented stereoscopically, one line to each eye, the fused lines will appear flat. Pastore (1964) showed that the introduction of a context which itself appears in depth induces opposite depth effects in the fused lines. The purpose of the present experiments was to test the generality of that induction effect (*IE*) to three stereograms involving other types of geometric figures.

Each stereogram consisted of two components: congruent critical figures and non-congruent contexts of vertical segments. In Fig. 1 the squares of stereograms *B* and *C* are congruent, as are the pairs of horizontal lines of stereogram *A*. The corresponding contexts of each stereogram are non-congruent; the noncongruency of the contexts of stereogram *C* is given by the differential displacements of the short lines above and below the squares. The length of a side of the squares, 13 mm, provides the scale for all other measurements. All subsequent statements of depth effects presuppose stereoscopic viewing. The anticipated depth effects and the results obtained are discussed together for each stereogram.

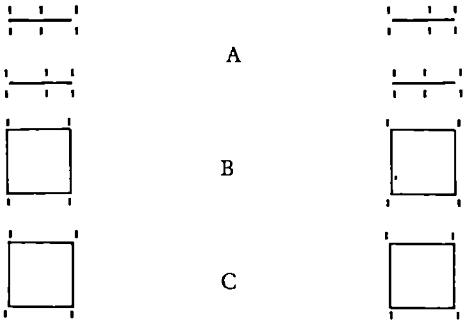


Fig. 1. Three stereograms in which the critical figures left and right are congruent (the pairs of parallel lines in *A*, and the squares in *B* and *C*), but in which the contexts of short vertical lines are non-congruent.

Method

Thirty undergraduates, who were able to achieve stereoscopic fusion, served as subjects. Before viewing a particular stereogram through a 'Keystone no. 46 Telebinocular' stereoscope, a subject was informed of the possible ways in which the critical figure might appear. For stereogram *A*, a subject was told that a horizontal line could appear flat, curved toward or away from him, or in some other and unspecified way; for stereograms *B* and *C*, a subject was informed that the squares could appear flat, rotated in space with one side closer to him than some other side, or in some other and unspecified way. The order of presentation of the three stereograms was varied from subject to subject.

Results and conclusions

Stereogram A. The contexts are such that the depth effects of the two pairs of end vertical segments are opposite to that of the middle pair of segments. With respect to the context of the upper line, the end pairs of segments appear forward and the middle pair appears distant. On the hypothesis that a depth effect in a given part of the context induces an opposite depth effect of points in its immediate vicinity, the ends of the line should appear distant and the centre of the line should appear forward. Since it had already been demonstrated that a line behaves as a perceptual unit, the entire line should appear curved with its centre toward the viewer. The context of the lower line should induce an opposite curvature, namely, with the centre away from the viewer. Six subjects reported these curvatures in both lines, twelve reported curvature in only one line.

Stereogram B. The right-hand segments appear distant and the left-hand segments appear forward. On the hypothesis of *IE* to a closed figure, the square should appear rotated in space with the right side closer than the left side. Twenty-five subjects perceived the square in this fashion.

Stereogram C. The upper segments appear forward and the lower segments appear distant. On the hypothesis of *IE* to a closed figure for this particular kind of context, the square should appear rotated in space with the lower side closer than the upper side. Thirteen subjects perceived this rotation.

These results support the generality of *IE*. However, introspective reports indicated that the *IE* differed in degree of pronouncedness. The perceived curvatures generally were 'moderate' or 'slight', whereas the rotations of the squares were pronounced. Moreover, it was easier to see the curvature in stereogram *A* with the centre of the line away from the viewer. Of the twelve subjects reporting curvature in only one line, two reported the centre of the upper line as forward, whereas ten reported the centre of the lower line as away. In this regard, it is to be noted that the rotation of the square was more readily observed in stereogram *B* than in stereogram *C*. Differences in the degree of pronouncedness of *IE*, then, occur between different stereograms and between different parts of the same stereogram.

It should be mentioned that *IE* disappeared for some subjects with continued inspection of the stereograms; on the other hand, a period of initial inspection was necessary for some subjects for the emergence of *IE*. In addition, all subjects who reported an *IE* were able to see the critical figures as perceptual units. Of those subjects who did not report the anticipated *IE*, there were only five anomalous reports all told; the remaining negative reports were of flatness.

The *IE* hypothesis is to be contrasted with other current approaches to stereoscopic depth effects, e.g. Howard & Templeton (1964). In further work on *IE*, it would be desirable to devise some procedure that will allow for the quantification of the reported depth effects. Moreover, it would be desirable to study the degree of *IE* as a function of the distance of the vertical segments from the critical figures and the degree of depth in the contexts.

REFERENCES

HOWARD, I. P. & TEMPLETON, W. B. (1964). The effect of steady fixation on the judgment of relative depth. *Quart. J. exp. Psychol.* **16**, 193–203.

PASTORE, N. (1964). Induction of a stereoscopic depth effect. *Science* **144**, 888.

(*Manuscript received* 18 *January* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 203–204
Printed in Great Britain

THE CONTRADICTION OF EPIPHENOMENALISM

By EDWIN A. LOCKE
American Institutes for Research, Silver Spring, Maryland

In the recent dispute in this journal between Keehn (1964) and Burt (1964) over consciousness and epiphenomenalism, Burt correctly points out that 'such a doctrine [epiphenomenalism] makes nonsense of the work of the educator, the psychotherapist, the moral reformer, the legislator, and the historian...' (p. 95). He might also have made one further point—that the doctrine also makes nonsense of the argument for epiphenomenalism and of anything else its proponents might say. Before showing why, let us be clear on the meaning of the doctrine.

Epiphenomenalism is the doctrine that all mental contents, for example, ideas, values, opinions, emotions, and the like, are an incidental by-product of physiological or physical processes and that they have no causal efficacy either in influencing behaviour or in influencing any other ideas the individual has. Let us for a moment assume this doctrine to be true and apply it to the behaviourist's own behaviour. Branden (1965) has pointed out that what the behaviourist is saying is that *the ideas and opinions which he* (*the behaviourist*) *has about the topic of epiphenomenalism have no relation whatever to the public statements* (*written or verbal*) *which he makes about it.* In other words, the fact that he *asserts* the doctrine to be true is in no way logically related to the fact that he *believes* it to be true. It is interesting to note that one never hears a behaviourist make such an assertion.

How then is the behaviourist to explain the fact that he *does believe* the theory to be true and at the same time *does behave* as if he believes it to be true, i.e. that he asserts it at all and that there is some empirical connexion between his beliefs and his public statements? As a consistent behaviourist he would have to assert that he was compelled *both* to think *and* to act in that way by stimulus forces outside his control, i.e. to invoke the doctrine of psychological determinism to explain both types of events.

But this only leads to a further contradiction which Branden (1963) has pointed out: if the behaviourist *could not help* believing and saying what he says, that is, if everything he thought and did were determined by forces outside his control, then how can he claim that what he is asserting constitutes knowledge or truth? If everything he does is determined by forces outside his control, the same must apply to his espousal of the theories of epiphenomenalism and psychological determinism. No proponent of psychological determinism can claim to have reached any conclusion on the basis of objective evidence which he is free to judge and evaluate; all he can assert is that he was forced to believe in and to assert his theories by his past conditioning, his childhood environment or the like. If men are not free to judge and to judge their judgements and to reach conclusions based on their assessment of the facts of reality, then no conclusion reached by anyone could make a valid claim to truth, including the theory of psychological determinism.

In sum, the argument for epiphenomenalism is untenable because (1) it leads to the rather absurd conclusion that the epiphenomenalist's own ideas have nothing whatever to do with what he says, and (2) to explain why he asserts the theory at all and why there is even a correlation between his ideas and his actions, the epiphenomenalist is forced to invoke the doctrine of psychological determinism, a doctrine which is incompatible with its own content and which would make all assertions of knowledge and truth meaningless.

REFERENCES

BRANDEN, N. (1963). The contradiction of determinism. *Objectivist Newsltr.* **2**, no. 17–20.
BRANDEN, N. (1965). *Basic Principles of Objectivist Psychology.* New York: Nathaniel Branden Institute.
BURT, C. (1964). Consciousness and behaviourism. A reply. *Brit. J. Psychol.* **55**, 93–6.
KEEHN, J. D. (1964). Consciousness and behaviourism. *Brit. J. Psychol.* **55**, 89–91.

(*Manuscript received* 5 *August* 1965)

Brit. J. Psychol. (1966), 57, 1 *and* 2, *pp.* 205–224
Printed in Great Britain

PUBLICATIONS REVIEWED

Compensatory Adaptations, Reflex Activity and the Brain. By E. A. ASRATIAN. Oxford: Pergamon Press. 1965. Pp. xv + 194. 80*s.*

Developed from a series of lectures delivered by Prof. Asratian in this country in 1957, this volume presents recent experimental findings obtained at the Moscow Institute of Higher Nervous Activity and Neurophysiology. It indicates clearly the direction of evolution of contemporary Russian neurophysiology. As would be expected, the tremendous research effort released by Pavlov's discovery of the conditioned reflex continues, and representative observations in this field are reported in the last three lectures. In the first three lectures, which constitute well over half the book, the name of Pavlov is also not infrequently invoked, but the spirit of the work, concerned in the main with spinal and other unconditioned reflexes, owes more to Sherrington, a debt which Professor Asratian generously acknowledges. Indeed, the recent Western departure from reflex physiology to microneurophysiology, with direct intracellular recording, seems to have had little effect on Asratian's work, which remains on strictly classical lines.

The first lecture presents experimental support for Sherrington's view of spinal block, the hypotonia and areflexia following acute spinal cord transection in its distal segment, as due to disconnexion from higher level facilitation, rather than to local mechanical irritation. The main changes are shown to take place not in the motoneurones but more proximally in the reflex arc. Restoration of function following neurological change is next considered. The integrity of the cerebral hemispheres is shown to be important for the development of effective compensatory reactions, and increasingly so the higher the organism on the phylogenetic scale—an interesting illustration of encephalization of function. The third lecture relates Pavlov's proposition that the unconditioned reflexes have cerebral representation to the observed weakening of homeostatic reflex responses and diminished resistance to extreme stress following decortication. The cortex is regarded as lending an adaptive variability to the relevant unconditioned reflexes. The unconditioned reflex arcs are regarded as branching at various levels in the neuraxis, including the highest.

In the last three lectures, the centre of attention shifts to the conditioned reflex. Just as Western experimental psychologists have tended in recent years to study *instrumental* conditioning under increasingly complicated conditions of reinforcement, so have the Russians investigated *classical* conditioning under similarly complex conditions. (Why doesn't someone study them both together?) Asratian's laboratory has been in the forefront of this movement, and many of his most important findings are here conveniently summarized.

The editor of the English translation, S. A. Corson, in a helpful footnote on p. 124, distinguishes three main types of experiment conducted by Asratian: 'switching' experiments, in which a conditioned stimulus is given one significance against one set of background stimuli and another in a different background—Western terminology would call this a 'conditional discrimination'; 'dual or binary conditioned reflexes', in which a single stimulus is reinforced simultaneously with both a reward and a punishment; and 'bidirectional conditioned connexions', which refers to Asratian's demonstration that, under certain conditions, not only does a conditioned stimulus elicit a conditioned response appropriate to the unconditioned stimulus, but also the unconditioned stimulus elicits a response appropriate to the conditioned stimulus.

All three types of experiments have produced results which should be of great interest to students of learning. Asratian's interpretation of the work on 'switching', reported mainly in chapter 4, depends heavily on the notion of tonic conditioned reflexes; this notion receives experimental support from an interesting experiment of Sakhiulina's, using the EEG in conjunction with the conditioned paw-flexion response (p. 139). The work on bidirectional conditioned connexions (chapter 5), should reopen the whole subject of 'backward conditioning'. And the research on binary conditioned reflexes contains important observations on the interactions between antagonistic drives and incentives (e.g. pp. 156 and 182–4) in classical conditioning. The latter findings should be of particular relevance to the efforts of such learning theorists as Miller, Spence and Amsel to account for some of the phenomena of instrumental learning in terms of classically conditioned 'goal' responses. Finally, Asratian describes how he has used all three forms of experiment for some ingenious attacks on the classical Pavlovian problem of the localization of inhibition; his results clearly exclude both the receptor and the effector elements of the conditioned reflex arc, leaving logical room only for his own view that the inhibitory block must develop somewhere in between.

The translation is very much better than is usual for this type of book, and Dr Corson is to be congratulated on his thorough, but unobtrusive, editing. We hope that this English publication of Asratian's work will be widely read. But why should such a slim volume cost the outrageous price of £4?

JEFFREY A. GRAY
M. KINSBOURNE

Animal Behaviour and Drug Action. Edited by HANNAH STEINBERG. Ciba Foundation Symposium. London: Churchill. 1964. Pp. xiii + 491. 67*s*. 6*d*.

The symposium published in this volume falls into two main parts. The first consists of the proceedings of a conference held at the Middlesex Hospital Medical School at which over twenty papers were presented, each being discussed in this context only briefly. The second part of the symposium, however, involved several largely unscripted discussions by a number of invited participants in which many of the topics raised in formal presentations were examined in greater detail, attention being given particularly to problems concerning the extrapolation—or, *interpolation* (Daws)—to man of drug effects recorded in animal experiments.

The publication of this symposium can be viewed as a significant addition to the growing literature of psychopharmacology and one can imagine that it will, of necessity, be consulted frequently by research workers in this area. Not only is the symposium noteworthy for a number of distinguished contributions, but it exemplifies clearly the inter disciplinary nature of drug-behaviour research, sampling such widely ranging topics as toxicology (Barnes,) electrophysiology (Bradley), behaviour genetics (Broadhurst), ethology (Chance and Silverman), operant conditioning (Cook; Dews), screening tests (Janssen), and drug-behaviour interactions recorded in the human experimental and clinical contexts (Irwin; Cole & Edwards; Hamilton).

The symposium reflects also the tentative and essentially developmental character of much current research pertaining to drug-behaviour interactions. Many of the authors are concerned to an appreciable degree with methodological issues; in particular, the papers by Weiskrantz and Russell illustrate some of the major difficulties associated with the interpretation of the behavioural effects of drugs and are outstanding in terms of their clarity. Prof. Russell deals at length with the problem of temporal interrelationships between the metabolic fate of drugs and their effects on behavioural processes, while Weiskrantz demonstrates the need to apply the 'double-dissociation' method of experimental analysis associated with ablation studies to the interpretation of drug effects.

The symposium included notable presentations by a number of overseas visitors including Miller ('The analysis of motivational effects illustrated by experiments on amylobarbitone'), Cook ('Effects of drugs on operant conditioning'), Dews ('Schedules of reinforcement'), Jarvik ('The influence of drugs upon memory'), Stein ('Amphetamine and neural reward mechanisms'), Bures ('The effect of physostigmine and atropine on the mechanism of learning'), and Russell ('The acetylcholine system'). The quality of these papers is exceptionally high and, as one might expect, many elegant experiments are described including a yoked-control self-injection study reported by Miller, the passive avoidance studies of Bures, and the intra-cranial self-stimulation/amphetamine experiments described by Stein, to mention but a few. From the theoretical point of view, a number of important generalizations about behaviour emerge which, no doubt, are being subjected to more intensive analysis.

It is not easy to give an overall impression of a symposium of this kind, in view of the diversity of the topics presented and the discussions that took place. Instead, several more specific comments can be made. First, one admires particularly the numerous contributions made by Prof. Neal Miller throughout the proceedings; in particular, one can cite the part he played in many of the discussions and especially the one concerning the effects of drugs on eating and drinking behaviour, in which he described the well-known experiments of Grossman, one of his students. Second, it is apparent that much careful editing was needed with respect to the unscripted discussions forming the second part of the volume. Often, recorded discussions of this kind, unless edited with care, appear to contain rather trivial and sometimes meaningless statements. This is not an impression one gains from the discussions appearing in this volume. On the contrary, this part of the book is as stimulating as the first.

Six main discussions took place representing the standpoints of biochemistry (McIlwain), electrophysiology (Bradley), neurophysiology (Feldberg), pharmacology (Gaddum), behaviour (Russell), and clinical work (Edwards) in which, apart from the principal speaker, many other participants contributed to a series of well-documented conversations. This part of the symposium afforded wide coverage of many topics of contemporary interest. For example, we learn a good deal from Bures about spreading cortical depression, Feldberg describes the effects of intraventricular drug administration on behavioural reactions obtained from the hypothalamus, and McIlwain talks at some length about the ionic and metabolic actions of drugs on isolated mammalian cerebral tissues.

Finally, two or three criticisms of the symposium can be made. The first of these relates to the fact that little, if any, attention is paid to the effects of drugs on perceptual processes, whether these be studied in animal or human experimental subjects; it is also true, more generally, that the symposium neglects to a surprising degree much work of importance in human *experimental* psychopharmacology which would not have been irrelevant to many of the topics raised. Thirdly, a number of statements can be found which are not challenged effectively. For example, it is simply untrue (p. 275) that operant conditioning techniques are poorly equipped to assess the effects of drugs on discrimination processes! However, these considerations apart, the symposium can be regarded as a successful undertaking which effectively brings together much current research in overlapping disciplines within the covers of a single volume. ROGER STRETCH

Deranged Memory: A Psychonomic Study of the Amnesic Syndrome. BY GEORGE A. TALLAND. New York and London: Academic Press. 1965. Pp. xi+356. 76*s*.

This excellent book reports a major piece of research and has the makings of a psychological classic. It contains, to quote the author, 'the lessons I have learned about the amnesic syndrome ...from observing and talking with patients, from experiments, from reading the reports and theoretical formulations of other students, and from pursuing my own speculations'. The inquiry focusses on eleven men and eighteen women who are in the chronic phase of Korsakoff's disease: they are closely alike as regards the aetiology and past history of their disease and also their social–cultural background: they are all hospital patients and were available for repeated examination over 6½ years. The first important question concerns the natural history of their condition. Precisely what are the psychological deficits and abilities of these patients? To answer this question a special laboratory was set up so that the patients could be given various tasks to perform under standardized conditions. Their performances, as a group, were then compared with those of two other groups, one selected to control for biological and social concommitants of habitual alcoholism, the other selected to control for any effects of prolonged hospitalization. These detailed group comparisons build up a complex and intriguing portrait of the Korsakoff group. The second important question concerns general psychology. Do the findings about these patients contribute anything to our views of normal functioning? Clearly, they do; and some admirable discussion is devoted to the implications.

The book is divided into five sections, each with its own distinctive merit. The first section outlines the goals and methods of the research. The second gives a general description of the amnesic patient and his everyday behaviour. The third section surveys the literature about such patients: it collates observations and theories which have been produced over the past hundred years and provides a valuable guide to the views of several neurologists and psychologists whose names are less well known than they deserve. The fourth section is the largest, for it contains details of the author's own experiments and results. The tasks used are many and various but, for presentation purposes, they are grouped under the following heads: verbal and motor skills, perception, intelligence and reasoning, learning and remembering, and the reproduction of narrative texts. Procedures and results are given in sufficient detail so that the interested reader can check his own theoretical hunches against the actual findings: for the less interested reader, there are useful summaries of groups of experiments. The final section contains the author's attempt at a rounded consideration of his extensive observations. This consideration is so well balanced and informed that it would be inappropriate for a reviewer to attempt a brief outline of the tentative conclusions. Only two points need to be made. The first concerns the amnesic syndrome itself. 'After examining my own observations and those of others, I am concluding it with speculations about a general activating function, and by proposing premature closure of this function as an answer to the psychological derangement under investigation.... They provide the material for a pluralistic explanation that is also subject to certain qualifications, so as to allow for the examples of efficient function retained by Korsakoff patients.' The second point is general to psychology. It is that many of the findings do not lend themselves to a ready fit with any of our current notions about behaviour and, indeed, that 'the need to revise our concepts of man's neuropsychology is precisely one of the lessons to be learned from the study of such diseases as the Korsakoff syndrome'.

In 1961 a preliminary version of this book was awarded the annual monograph prize of the American Academy of Arts and Sciences in the biological and physical sciences. Now that the book is published it is evident that this award was fully merited. The author has tackled important psychological problems with a refreshing blend of scholarship, clinical sensitivity, experimental sophistication, and common sense. If he has not fully resolved these problems, it is because of their intricacy rather than any incompetence of his part. What he certainly has done is to provide clinical and experimental psychologists with a model for careful inquiry and clear reporting. For the general psychologist, he has furnished many fresh facts of a challenging kind and an illuminatingly broad-minded attempt to co-ordinate these and other facts into a coherent pattern.

I. M. L. HUNTER

Handbook of Mathematical Psychology, vol. 3. Edited by R. D. LUCE, R. R. BUSH and E. GALANTER. New York: Wiley. 1965. Pp. ix+537. 90*s*.

The high standards set by the first two volumes in this series have been maintained in the third, both in respect of what is presented to and what is demanded from the reader. However, this final volume covers an astonishingly wide range of subject-matter, from sensory physiology to the theory of games, and although this is stimulating, it may result in the volume not being such a good buy on its own as was either of its predecessors.

There is an excellent chapter on auditory characteristics, in which the author (Zwislocki) demonstrates impressively the value of biophysics to an understanding of sensory psychology. The treatment of basic harmonic functions and electrical–mechanical–acoustical analogies should be of interest to all those concerned with dynamic phenomena, acoustic and otherwise. The chapter on visual problems is disappointing by contrast. Some selection from such a wide field is inevitable, but with three-quarters of the chapter devoted to a single topic (colour-vision), treatment of other problems is scant.

The longest chapter in the volume is an account of utility and preference theory. The treatment is unremittingly theoretical—in many places it far out-runs empirical observation, and in some places even the possibility of such observation. Some may find this disturbing. The potentialities of mathematical inference, so apparent elsewhere in the *Handbook*, here take on an *a priorism* reminiscent almost of Herbart. The reader is warned of the non-empirical theories of economists and statisticians—it would seem that psychologists are not themselves innocent of such theories.

There are two good chapters on aspects of learning-theory, which might more sensibly have been included in volume 2, which contains the main treatment of this subject. Two chapters on mathematics complete the volume; one an intermediate-level treatment of stochastic processes, the other an introduction to functional equations, a subject previously accessible only in very advanced texts but which underlies much of the material in the *Handbook*. One might have wished to find among this diversity of topics some treatment of neurone behaviour and network theory, but coverage of this subject is limited to a brief mention in volume 1.

It is difficult to assess the usefulness of the complete *Handbook* in this country, where to take full advantage of it the reader requires mathematics and psychology each to degree level—a luxury afforded to only a few. Rightly perhaps, the work is uncompromising toward the mathematically inexperienced reader, who without some acquaintance with the introductory bibliography in volume 1 (updated, incidentally in, volume 3) would be lost within the first three pages of any chapter. However, this *Handbook* is important evidence that mathematics (as distinct from statistics) has come to stay in psychology. If so, it must eventually receive proper representation in the teaching of psychology—the prospect of 'cook-book mathematics' is not an appetizing one. For the moment, the material is not beyond the capabilities of energetic postgraduates, to whom this *Handbook* should be a great help in understanding modern research. They should be encouraged to buy it—it needs living with.

D. H. TAYLOR

An Introduction to Psychological Statistics. BY PHILIP H. DUBOIS. New York: Harper and Row. 1965. Pp. xi+530. 60*s*.

Prof. DuBois assumes that the best introduction to psychological statistics is through techniques applicable to individual differences, on the grounds that most psychologists are applied psychologists whose main use of statistics is in interpretation of test results and prediction of individual attainment. By emphasizing descriptive rather than inferential statistics, he is able not only to provide an unusually detailed and comprehensive introduction to measures of association and correlation and their uses in forecasting behaviour, but to add an excellent introduction to factor analysis and a full chapter on the logic and use of matrix algebra.

Throughout, the main emphasis is practical. Fully detailed manual and machine computational techniques, including many alternative and new formulae, are given, carefully chosen practical exercises conclude each chapter and a separate instructor's manual provides full numerical working for each exercise and objective test items for checking progress. But, despite his preoccupation with techniques, he does not fail to point out the logic of basic formulae or the conditions of their use. Indeed, in this respect his exposition of multiple correlation and his discussion of the relations among basic concepts and among probability distributions are among the book's best features.

But, perhaps inevitably, this approach leads to pedagogic difficulties. Apart from the obvious difficulty of seeing the logical wood in the dense forest of computational trees, the main source of possible confusion is the separation of instruction in the descriptive uses of statistics from discussion of their uses in inference and decision. Thus the first ten chapters are largely devoted to spelling out definitions and computational procedures. Discussions of probability, sampling distributions and statistical inference, including a curiously unenthusiastic introduction to analysis of variance, are then squeezed into four chapters, followed by an immediate return to descriptive statistics via test construction and factor analysis. A short final chapter describes a few basic non-parametric statistics.

To the reviewer, three weaknesses result. First, no attempt is made to relate statistical inference and analysis of variance to design of experiment. Indeed, the short discussion of statistical control of experimental variables mentions only partial correlation as an analytic technique. Nor is mention made of conditional probability, Bayesian strategies or statistical decision theory (which this reviewer would have thought crucial to discussion of the use of test data in individual diagnosis, selection and prediction). In fact, apart from the brief excursion into non-parametric statistics, the text owes little to developments of the last decade or more.

Secondly, not only is the account of each descriptive statistic separated from consideration of its reliability and significance, but difficult theoretical concepts are perforce introduced and used before they are formally defined and discussed. Inevitably they become fossilized in misleadingly concrete assertions. Typical is the assertion, in chapter 3 (which attempts to teach the uses of chi-square nine chapters before formal discussion of probability and sampling distributions), that 'it can be shown that "change" or, better, "probability" operates according to definite principles' (p. 57). The author justifies this approach on the grounds that it reduces 'symbol shock'. If a choice must be made between 'symbol shock' and 'conceptual paralysis' this reviewer hopes he is not alone in preferring the former.

The first two weaknesses illustrate the third. No attempt is made to foster critical thinking—indeed, the emphasis on descriptive techniques appears to discourage it.

Thus, although clearly not for psychologists seeking an up-to-date introduction to statistical inference, experimental design or statistical decision, the book deserves recommendation as a good practical introduction to correlational analysis and kindred techniques. But, unless supported by adequate teaching, this reviewer would be unhappy to think that it was any psychologist's only introduction to contemporary statistics.

A. B. ROYSE

Scientific Psychology. Edited by BENJAMIN B. WOLMAN and ERNEST NAGEL. London: Basic Books. 1965. Pp. xv + 620. 70*s*.

The title of this important collection of mainly original essays is unfortunate, for the book presents a timely attempt to bridge the gap between the Minnesota Studies in the Philosophy of Science on the one hand and the ubiquitous 'Contemporary Theories' type of publication such as Marx and Hillix on the other. It is indeed, as the publisher's blurb claims, something of 'a pioneering attempt to effect a rapprochement between scientific psychology and philosophy of science'. It might have been, in truth, equally misleadingly entitled *Philosophical Psychology*.

With his powerful analysis of the methodology of psychology and the social sciences in his recent *Structure of Science* as a background, Ernest Nagel has gained a special place among logicians of scientific method and his consultant editorship of this volume has attracted an impressively international assembly of philosophers of science to counterbalance the baker's dozen of psychologists who, with the solitary distinguished exception of Piaget, have been exclusively, no doubt for convenience, recruited from the U.S.A. A leavening of mathematicians, psychiatrists and educationists—again in the main American—make up the full team of thirty contributors. Included are interesting chapters by W. Horsley Gantt, Mary Henle, Otto Klineberg, Lawrence Kubie, Charles Morris, Benoit Mandelbrot, R. S. Peters, Karl Pribram, Anatol Rapoport, Gregory Razran, Michael Scriven and G. J. Warnock, writing predictably on their special interests. The contributors vary as one expects in length and importance and there is small correlation between these two measures.

Nagel's own contribution is a most disappointingly small three-page comment without bibliography and one could not suspect from it that he had been deeply interested in the complexities of the problem. The 'paper' might have been considered as a short introduction if Wolman had not written this himself besides a rather opaque piece about something called 'Monistic transitionism' thus concluding: 'Transitionism agrees that what can be reduced, should be reduced. Certainly there is little reason to force in psychology a radical, general, here-and-now reduction at the empirical level. On a formal logical level, transitionism seeks to develop a set of constructs, in psychology and other sciences that deal with living organisms, that would enable us to present mental phenomena not as logical derivative but as a *continuation* of the other organic processes. A programmatic and imaginary formula would, more or less, represent O (organic phenomena) that can in certain conditions of k (conditions of transition) turn into M (mental phenomena) and vice versa. Thus (1) $O_k \rightarrow M$ and (2) $M_{kl} \rightarrow 0$.' GEORGE WESTBY

Inside the Black Room: Studies in Sensory Deprivation. By JACK VERNON. London: Souvenir Press. 1965. Pp. xvii + 203. No price given.

The informal, non-technical style of writing makes this a very readable book, quite suitable for the layman. For the psychologist also, who knows little about sensory deprivation it will provide a painless entry to the subject and one which should not take more than an evening to read. He should emerge, not unduly fatigued, with a clear idea of the procedures and difficulties of sensory deprivation. Unfortunately, he may also emerge with some misapprehensions about what is known in the field. As often happens with informal presentation, the line between fact and speculation is not clear. Often it is difficult to tell which even of the author's own results are reliable and which are not. Conventional levels of significance are claimed only sporadically—and in one place (the reaction-time studies) even the figures are inconsistent. As a presentation of scientific data, then, this monograph is unsatisfactory and likely to irritate the serious student. A further source of frustration for the latter is the lack of any formal references.

The experiments themselves deal with the effects of sensory deprivation of up to four days on a variety of human abilities both mental and 'physical'. A good description is given of the various hallucinations which may occur and Vernon makes the important point that his subjects showed much less of this activity than did those of the earlier McGill studies. Other effects considered are those on learning tasks of varying complexity, estimation of time-intervals, pain threshold, pursuit rotor tracking, reaction time, problem-solving and suggestibility. Reaction time data receive rather more space than they deserve in that they are presented twice almost identically in different parts of the book. Finally, Vernon stresses a hitherto neglected point, namely that many of his subjects commonly reported sleeping during much of the first day of deprivation but that later they found sleep progressively more difficult. This may mean that sensory deprivation is in some ways the opposite of sleep deprivation: a satiation of sleep, or at least, of a sleep-like level of stimulation. The author does not make this point, however, and in general devotes little attention to the theoretical side of his subject. He mentions its possible relevance to developmental aspects of perception and neurologizes briefly and dangerously in terms of its possible links with brain-stem reticular function. Finally he underlines what is perhaps the most important point, that man's inability to tolerate a chronically reduced sensory input emphasizes the body's need for a certain degree of sensory priming.

To sum up, this is a readable and accurate description of the techniques of sensory deprivation. As a review and an evaluation of the knowledge in the field, and as a scientific report of the author's own data, it is less satisfactory.

R. T. WILKINSON

Psycholinguistics: A Survey of Theory and Research Problems. Edited by CHARLES E. OSGOOD and THOMAS A. SEBEOK. With *A Survey of Psycholinguistic Research, 1954–64* by A. RICHARD DIEBOLD, and *The Psycholinguist* by GEORGE A. MILLER. Bloomington: Indiana University Press. 1965. Pp. xii + 307. $4.75.

This paper-back contains the equivalent of a 200-page book, an 85-page monograph and a 15-page article which, respectively, illumine the past, present and future of psycholinguistics.

The book within the book is a reissue of the out-of-print work edited by Osgood and Sebeok, which first appeared in 1954 as a supplement to the *Journal of Abnormal and Social Psychology*. This was an early, and influential, attempt by a group of psychologists and linguists to map out the area of common concern that has become known as psycholinguistics. It is still worth reading, and not only as a historical document. Its analyses of basic theory and problems (for example, Saporta's discussions on pp. 60–5 and 77–83 of psycholinguistic units of analysis) and its frequent suggestions for research merit attention. On the other hand, the structuring of the field, though seemingly systematic, probably makes even less sense now than it did eleven years ago, and the absence of well-organized references is annoying.

Diebold in his review of research from 1954 to 1964 ignores the framework used by Osgood and Sebeok in favour of the arbitrary but intelligible list of headings used by Saporta in his excellent *Psycholinguistics: A Book of Readings*. These topic headings were: The nature and function of language; Approaches to the study of language; Speech perception; The sequential organization of linguistic events: Language acquisition, bilingualism, and language change; Pathologies of linguistic behaviour; Linguistic relativity. To these Diebold adds a brief introduction to psycholinguistics and an 'Other topics' section, in which he concentrates on animal communication and human non-verbal communication. Thus most psycholinguistic problems are covered in the review. The major omissions are reading, writing and the acquisition of foreign languages, which have been areas of unfortunately marginal status in psycholinguistics. This monograph is likely to prove extremely informative for psychologists interested in language. It covers a mountain of material from a variety of disciplines. In all, there are 450 references of which four-fifths have, in fact, appeared from 1954 onwards. Furthermore, on occasion Diebold has allowed himself enough room for discussion to cause the enthusiast, as he greedily files away the references, to pause and think.

The volume ends with the essay *The Psycholinguists*, in which Miller presents one interesting view of major problems which await solutions. The reader, however, should not allow the pleasantness of style to lull him into accepting uncritically the controversial content; for example, the treatment of language development in the child.

This book should be a useful acquisition for those already involved in psycholinguistics and, it could be of value to anyone who is prepared to put some effort into finding out what psycholinguistics is about.

COLIN FRASER

Psychology of the Child. By ROBERT I. WATSON. New York: Wiley. (2nd edn.) 1965. Pp. 635. 57*s*.

This is the second edition of a comprehensive book which, originally published in 1959, has doubtless already proved its worth to both undergraduate and postgraduate students taking courses in child psychology. The improvement in style in this edition makes the book more readable than before. There has also been a slight rearrangement of the contents. The main divisions of the volume are still Principles of Development, Infancy (the first two years of life), Early Childhood (up to five years of age), and Later Childhood (up to the thirteenth year); but within these divisions, there are new sub-headings. Each period is now discussed in terms of Individual Development, Parental Influences, Behaviour Tendencies, Self and Peer Relationships and Psychoanalytic Contentions. Professor Watson has chosen this arrangement rather than the usual format of text-books in child psychology in order to give a clearer picture of the whole child. The new arrangement is not entirely successful as it leads to some repetition and overlapping of content in places and, at times, to the premature interruption of an argument.

However, this is a minor criticism of a book which can be enthusiastically recommended. Its main value lies in the author's selection and treatment of research studies, now brought up to date. Many writers of similar text-books are content with a brief mention of a large number of studies, encouraging a superficial approach, but Watson certainly achieves his aim of giving enough detail of the researches considered to allow the reader to see the significance of each study in its particular context. Not only does he summarize research evidence with commendable clarity, but he makes apt critical comments and attempts either to reconcile apparently conflicting findings from different studies or to explain why the findings were at variance. As is usual in American text-books, relatively little attention is given to British research studies, but these are not altogether ignored, and the British reader will find it useful to have so many American studies of importance, mostly scattered in a variety of journals, surveyed in such detail in a single volume.

The introductory sections, after giving the reader an historical background and describing scientific methods of child study, clarify some difficult concepts. In the main sections of the book, physical development is discussed only sketchily, but there is very detailed treatment of cognitive, language, and social and emotional development. Piaget's work is dealt with more fully than in the first edition, and the new chapters on parental influences provide a synthesis of many stimulating studies. The author takes a broad view of child psychology, giving attention to findings from clinical as well as from laboratory studies. He is particularly impressed by those psychologists who have tried to build a theory of personality development on the basis of learning theory, and gives much space to the work, for example, of Dollard and Miller on social learning, Robert Sears and his colleagues on frustration–aggression and patterns of child rearing, and Whiting and Child on socialization.

The chapters on psychological disturbances in childhood which formed part of the first edition have not been given a place in the present edition, probably because the book was felt to be long enough as it is. The book thus deals almost entirely with 'normal' child development, though a number of the studies quoted throw light on the aetiology of disturbed behaviour in childhood.

The suggestions for further reading and the lists of references which follow the majority of the chapters enhance the value of the book, in which both statements of facts and interpretations of theories are consistently lucid.

MAURICE CHAZAN

The Adolescent Spurt in Mental Growth. By BENGT-OLOV LJUNG. Stockholm: Almqvist and Wiskell. 1965. Stockholm Studies in Educational Psychology 8. Pp. xxx + 350. Sw. Kr. 48.

The balance of evidence from previous research, mostly longitudinal growth studies, is *against* the existence of any mental growth spurt similar to the spurt in height at puberty. Dr Ljung's findings differ. 'The results provide strong evidence of a mental growth spurt for girls aged 12–13. It is less evident whether a corresponding spurt exists for boys aged 14–15; if so, it is less pronounced than for girls' (p. xxx).

The data in this study were derived from cross-sectional-samples of the Swedish population. Dr Ljung argues that a cross-sectional approach is preferable to a longitudinal study. Since girls reach puberty before boys, his method of testing for a mental spurt at adolescence is to examine sex differences in score. Within each year-group, he calculates the regression of score on age, month by month, for boys and girls separately. The slope of the regression lines and their curvature when quadratic equations are fitted are evidence for or against the presence of a spurt or a plateau at each age.

'Mental development' is assessed from the standardized school tests (reading, writing, spelling, punctuation, arithmetic and others) applied nationally in Sweden. By factor analysis of the tests, factor scores are derived; these, normalized and transformed to *t*-scores, provide his basic measures of mental development. The computation is done separately for boys and girls at ages 9, 10, 12 and 14. Regression on age was calculated on samples of over 1000 of each sex at each age (that is, about 100 per month) by selecting children born on a specific day or days in each month of the year. Bright youngsters who had been accelerated through school and dull older pupils who had fallen behind were omitted.

The discussion of results is extremely thorough: for example, nine sources of bias in sampling are discussed in full. The statistical treatment is highly sophisticated, as can be seen from a glance at the 341 tables. There is an honest presentation of results, whether these support the hypotheses or not. However, three substantial doubts may be voiced.

First, it is questionable whether dependable evidence can be got by fitting quadratic equations and noting the slope of regression lines, where small fluctuations can produce substantial changes. (For example, the slope can be reversed by omitting January or December subjects). The author concedes that 'the hypothesis of linear mental growth trend lines cannot be rejected' (p. 230). In this study on the adolescent spurt, there are no data at all on the onset of adolescence: the whole argument rests on the gradients of the regression lines for boys and girls. Second, there is a danger that this finding may be quoted to justify special consideration for late (or early) physical developers. The actual size of the spurt (if it does exist) is tiny, compared with the range of individual differences. The author argues that his experimental design underestimates the spurt. He also admits that 'the influence of environment and learning has a smoothing-out effect on the relatively small differences in mental maturation' (p. 229). Thirdly, the regression of age on score depends on test content as well as on the maturation of the pupils. The use of factor scores does not mean that an innate growth is being assessed.

Whatever the final verdict, this work will be widely read and quoted by those concerned with development in adolescence, and it deserves attention because of its comprehensive discussion of growth studies and because of the new approach made possible by the extensive Swedish testing programmes.

JOHN NISBET

Innovation and Research in Education. By MICHAEL YOUNG. London: Routlege and Kegan Paul. 1965. Pp. 184. 25*s*.

Michael Young has written a very valuable book for the right readership, but that readership is primarily educational administrators and teachers, not psychologists. If, however, his aims are to be implemented, psychologists will necessarily need to be involved.

The main aim of the book is to dovetail changes in educational practice with ongoing validational research in schools. This might seem to be an obvious procedure but Young argues convincingly that educational practice has been governed more by changing fashions than by careful inquiry and validation. In fact, he is extremely critical of the role of 'theory' in education, claiming that it 'has not so much freed education from the hold of fashion as added yet further strength to its strands'.

He is sensitive to the need for questioning many of the basic assumptions underlying the present educational system, and this includes the widely held assumption that education itself is necessarily a 'good thing'. Over and above problems of curriculum change Young picks out for further investigation the twin themes of social class, and of home and school relationships. This emphasis is certainly not misplaced.

In discussing the implementation of research schemes in education, he is fully aware of the pitfalls which have to be circumvented, and bearing this in mind, he opts for the setting up of experimental schools rather than broadly based projects. He is probably right in this preference.

Carrying out of the research should be in the hands of institutes covering different aspects of educational practice, and these institutes might well be based on existing departments of education.

There are some interesting omissions. There is no mention of seeking from the teachers and the consumers of education what they feel needs to be done, nor is there much mention of the role of the educational psychologist in formulating questions and taking part in research. These workers might never exist. It is also a pity that he uses the expression 'I.Q. tests' instead of the straightforward 'intelligence test'. On that score, however, he is in the same company as many who should know better.

These shortcomings apart, Young's book needed to be written, it needs to be widely read, and some psychologists might feel that the problems which he has posed will present a worth-while challenge.

A. T. RAVENETTE

Deprivation and Education. By M. L. KELLMER PRINGLE. London: Longmans. 1965. Pp. 307. 42*s*. 6*d*.

Dr Kellmer Pringle and some of her senior students at the Department of Child Study, University of Birmingham, have studied for some years the effect of residential care and deprivation of normal home life upon the intellectual, educational and emotional development of children. The present book comprises several articles, previously published in learned journals, together with articles on the emotional needs of children and on the practical prevention of the worst effects of deprivation.

The studies cover such questions as the range of intelligence and the incidence of emotional disturbance among children in residential care compared to an ordinary school population. They examine questions such as the effects of early as opposed to later separation from the mother and family, the effect of length of stay in care and the effect of continuing as compared to complete loss of contact with the family. As one study arises from another, the evidence emerges that, as workers in the field know well, the incidence of maladjustment among children in residential care is high, that the 'severely maladjusted' tend to be those who are rejected and unwanted by their parents in contrast to the 'notably stable' who tend to receive continuing parental interest and contact, and, perhaps more significant still, that many deprived children are markedly backward in speech and language development and retarded in reading.

Dr Kellmer Pringle discusses the emotional and relationship needs of children and she offers guidance to child care and nursing staff on the steps which should be taken to provide for the appropriate emotional and intellectual development of young children in care. She stresses the benefit, to those children who have little parental contact, of dependable 'uncle and aunt' contacts outside the Home and of small classes and remedial teaching for the many who are unsettled and retarded.

One could criticize this book. It is uneven in style in that it combines rigorous research articles with more discoursive chapters, and there is a good deal of repetition. However, workers in the general field of child care will find much in this book which will help to clarify their ideas. The book ends with a valuable chapter in which a broad model of primary prevention is described. Clearly professional effort of many kinds must increasingly be directed to preventing the breakdown of families and to providing appropriate supportive help during temporary family crises. Dr Kellmer Pringle makes an important contribution to the discussions on this crucial aspect of child care.

MARIA C. ROE

Maladjusted Boys By OTTO L. SHAW. London: George Allen and Unwin. 1965. Pp. 168. 28*s*.

To those privileged to have seen the disconcerting, stimulating, iconoclastic originality of Otto Shaw in full therapeutic action with the most difficult of disturbed young people at Red Hill School in East Sutton, Kent, this book must be at once a delight and something of a disappointment.

It is a delight in that it tells for the first time, and authoritatively, the full story of an unusually impressive piece of pioneering in a field in which many have failed because the qualities required are so varied. In the establishment of a fully efficient residential community providing psychological treatment, which will recommend itself both to hospitals and to local-authority child guidance clinics, the combined skills of the entrepreneur, the administrator and the psychotherapist are all demanded, besides the necessary high regard for educational values of a school principal. Shaw has all these in good measure, and, where many others have fallen short of these high demands, he has dramatically succeeded. Some have suggested that his task has been an easier one than A. S. Neill's, in that he learned much from the earlier pioneer and had a little more sympathy with the basic ideals of our established society (e.g. with the values of orthodox religion), but Shaw's determination to tackle the highly intelligent group of the maladjusted presented him with very special problems in both staffing, organization and therapy, though of course there were entries on the credit side of the balance sheet. In a school where the *average* I.Q. is well above that required to get an Honours degree at University, Shaw has not made the mistake of providing an intellectual precinct lacking in the necessary stimulation to activity within the community. The school, as the book makes clear, is a hive of spontaneous activity in the arts: it should be said that the fees to be paid by local authorities for their problem-children's education and treatment are as high as those of a top public school.

The measure of disappointment will perhaps not be felt by those who do not know Shaw well. But the reader is really brought only very occasionally to glimpse the 'real Shaw'. I picture his Jaguar arriving down the steep drive to the school before nine in the morning with a convoy of small boys before and behind excitedly screaming 'Shaw's here!' This ritual welcome bespeaks a magic of personality which the book does not reveal—perhaps it should have been written by a biographer. Without this picture of the personality the reader will be puzzled to explain the skills which have been the secret of his obvious success with children, for he is certainly not a theoretician bound to an orthodoxy. Those who know him well will make their intuitive guesses and list a number of gifts and some obviously painstakingly acquired techniques, some of which emerge in the story; but to enlarge upon them here might be misleading to those with no experience of the human needs arising in situations of residential psychotherapy.

It is his achievement that Red Hill School has become a unique sociotherapeutic community, ideally suited to the intelligent, imaginative, often precociously sex-involved boy in the pre-adolescent and adolescent stages of development. The culture-pattern of this miniature society is something to which the boys can feel a sense of loyalty which is revealed in many instances in the book. The structure of human relations allows sufficient freedom of manœuvre during the critical, often disturbing, stages of psychotherapy, but at the same time it provides continuing stimuli to social and personal integration through creative activity. The contribution of his able colleague Ivor Holland to this achievement is well shown in the appendix he has written.

The pity is that there is no similar provision in our society for girls. Shaw's attempt to maintain the co-educational principle, despite his tenacity of purpose, came to an end soon after the war. The very much smaller number of girls of high intelligence referred for treatment and the special problem arising from their minority position in the school made it inevitable that Red Hill School should become a boys' school.

Can one imagine a female pioneer with the abilities to tackle in similar fashion the development of a self-governing psychotherapeutic and educational community for the perhaps even more difficult group of highly intelligent, precocious and maladjusted pre-adolescent and adolescent girls? It is indeed difficult, but then it would have been impossible to imagine an Otto Shaw.

GEORGE WESTBY

Modern Perspectives in Child Psychiatry. Edited by JOHN G. HOWELLS. Edinburgh and London: Oliver and Boyd. 1965. Pp. xvi+595. 105*s*.

Child psychiatry suffers from a shortage of text-books so that the appearance of any serious contribution in this direction is particularly welcome, especially when the authors are writing from their experience of work in this country. This book's main weakness is unevenness of quality and level, but it is likely nevertheless to be widely consulted, not only of necessity by those professionally concerned, but also by other disciplines.

The book is divided into two parts—the first concerned with 'The scientific basis of child psychiatry' and the second with clinical matters. Outstanding in the first section are contributions by Gwynne Jones on research methodology, Woodward on Piaget's theory and Eysenck and Rachman—at great length—on learning theory in relation to child psychiatry. The first section also contains useful chapters on genetics, ethology, perception, thinking and psychopathology and an important contribution by Stephen and Robertson on the relevance of certain concepts of child development to handicapped children.

The second section on clinical problems is somewhat patchy. Some chapters adopt the descriptive and taxonomic style of the standard text-book. Others are written as high-level critical reviews of the literature, while space is also found for provocative and stimulating discussions, as, for example, that on psychosomatics by Pinkerton, however much one may disagree with the subjectivity with which he concludes. Several chapters are devoted to unusual but important clinical problems about which there is little guidance in the literature—Hilda Lewis on adoption, Connell on suicidal attempts and a lucid exposition by Pond on the 'brain-damaged' and epileptic child.

Dr Howells himself is a firm believer in a hospital-based concept of child psychiatry and has scant sympathy for the traditional child guidance trinity of psychiatrist, psychologist and psychiatric social worker. Although the somewhat dogmatic tone of his own contribution on organizational aspects may offend a few professional sensibilities, the book as a whole provides a balanced presentation. A certain unevenness in comprehensiveness of approach and some inconsistency in deciding on the level of the audience are apparent, but perhaps this is unavoidable when twenty-four authors are involved in an enterprise of this kind. There are those who regret the passing of the one-man text-book, but the sacrifice of unity and consistency may be justified, as in this volume, by the excellence of individual chapters.

PETER MITTLER

Infantile Autism: The Syndrome and its Implication for a Neural Theory of Behaviour. By B. Rimland. London: Methuen. 1965. Pp. xi+282. 36*s*.

For someone like this reviewer, who thinks that progress in abnormal psychology is most likely to be made by using the techniques and methods of experimental psychology, a book like this presents some difficulties. The author puts forward some useful and interesting ideas. He suggests, for instance, that insufficient connexions between percepts and memories may become established in autistic children, thus making learning impossible. Rimland thinks that this is not so in schizophrenic children. But important as a precise sub-diagnosis of this obviously heterogeneous group of psychotic children would be, as treated by Rimland it rather implies that it is a distinction to belong to the exclusive autistic club as compared with the more mundane schizophrenic one. The author's questionnaire for establishing a sub-diagnosis lacks any attempt at statistical verification. The underlying attitude is reminiscent of the era in mental deficiency when people like Strauss and Lehtinen were determined to distinguish between brain-injured and non-brain-injured defectives. These terms were so vague as to remain practically meaningless and the tests based on them were subsequently found to be largely invalid. Psychotic children, whether called autistic or schizophrenic, share many behaviour characteristics on which research that aims at treatment and education should be based.

It is pertinent to propose, as Rimland does, that the so-called emotional withdrawal of autistic children, much stressed by clinicians, may in fact be due to cognitive rather than to emotive deficits. It is to be welcomed that the author rejects, as unsupported by evidence, theories about a psychogenic origin of the illness, which links the syndrome to parental attitudes, and to the parents' treatment of the child. But Rimland's more biologically orientated approach, though laudable as an attempt, lacks rigour and leads to much naïve neurologizing. His hypothesis about a genetically determined impairment of the reticular formation of the brainstem in infantile autism is vague and speculative. The author tries to hitch his ideas to some of those psychological bandwaggons which at this precise moment are about to turn the corner into oblivion. The initial reports from Hebb's laboratory, for instance, comparing the effects of sensory deprivation with some manifestations in mental illness, had to be considerably modified when investigated in controlled experimental conditions.

The style of the book, often more reminiscent of science fiction than sober scientific reporting, would be more acceptable if the author's theories were not mainly based on the same kind of clinical reporting, mostly uncontrolled and unchecked for reliability and validity, which unfortunately characterize the literature in this field. Thus Rimland asks: 'What would be the behaviour of a child whose reticular formation were badly impaired? Is this the secret of the veil of autism?'

It is to be hoped that he may eventually be tempted to answer this and other questions raised on the basis of more reliably obtained factual data than that on which he bases his theories in the present book.

BEATE HERMELIN

The Causes and Cures of Neurosis. By H. J. Eysenck and S. Rachman. London: Routledge and Kegan Paul. 1965. Pp. xii + 318. 28s.

In many ways this book is disappointing. Forty-five years after little Albert, one might hope for an objective evaluation of conflicting theories about why some people become neurotic and others do not, or a discussion of the gaps between learning theories, the laws of learning, and what behaviour therapists do. Failing this, one might hope for a cook-book with detailed therapeutic recipes and hints on how to manipulate inhibition and excitation with drugs and split-second timing in relation to personality measures. Instead, we have the Eysenckian theory again, detailed anecdotal discussions of two or three cases and a useful but repetitious annotated bibliography. Much of this information has been readily available for several years, and the theory needs discussion rather than simple presentation. The style is rather stodgy and the carving knife is only occasionally flourished with accustomed panache (e.g. against Oswald, p. 163, and Cooper, p. 250).

The first four chapters and the last present the author's general position on personality theory and its implications for the understanding of neurotic disorders. Behaviour therapy and psychotherapy are both accorded considerably more self-consistency than seems reasonable: thus, the realistic, messy state of affairs appears clearly enough in the chapters on different syndromes but this is not allowed to blur the black and white clarity of the theory. Neurotic disorders are considered to be simply maladaptive habits on the one hand or deficits of learning on the other. In either case the symptomatology of the particular individual is a function of his conditionability, autonomic lability and accidental environmental factors. Dysthymic symptoms are conditioned sympathetic reactions which appear to be built up into neurotic disorders (instead of being extinguished) because the individual finds that he can in some way avoid the anxiety produced by the conditioned stimulus and hence shuns extinguishing situations before the anxiety can be unlearnt. This sort of disorder is contrasted with those where the individual is poor at this type of learning and so does not develop the socially desired avoidance habits or learns some socially deviant but personally satisfying habits. The final chapter brings out the point that in the first case the therapist is trying to help along the normal extinction process and in the second case he is working antagonistically to it. This is why, it is claimed, spontaneous recovery and relapse rates are markedly different for the two groups of disorder. This leads to what is perhaps the most important point in the book: we should not wonder why therapy works or fails but why spontaneous remission occurs in the well-known two-thirds of cases. Certainly it would seem that the theoretical position put forward in the book can cope quite comfortably with spontaneous remission, and less adequately with therapeutic failure, but it is not clear that alternative theories would be in any great difficulty.

While using the avoidance conditioning model to explain symptom formation, the possibility that symptoms may serve a useful purpose seems to be denied. It is not surely necessary to postulate unconscious mechanisms to entertain the idea that superficially maladaptive habits may have once got the organism out of trouble (and not merely been caused by fright) and may indeed still remain adaptive long afterwards.

Another oddity of the book is the assumption that there is some single entity called 'behaviour therapy' and that this is closely related to another entity called 'learning theory' in a systematic way: this is clearly contradicted by the detailed discussions of particular techniques in the practical chapters which seem far indeed from Pavlovian, Hullian or (even) Eysenckian learning theories.

The chapter on the results of behaviour therapy deserves special discussion since the data on efficacy have not been so readily accessible before. It begins with an attack on clinicians like Wolpe who have not had time to collect adequate controls. This is reasonable. Despite appropriate cautions, we are treated to Wolpe's 90 per cent cured or greatly improved as a major advance on the famous two-thirds spontaneous remissions. We also hear of Lazarus's 78 per cent (62 per cent of the more serious cases) on the next page, but no reference is made to the magic figure this time and it is suggested that Wolpe did better because he tried harder. There follow studies by Hussain and by Burnett and Ryan, with sufficient detail to show gross deficiencies in the work. It is suggested that this uncontrolled data provides evidence for hypotheses that behaviour therapy is effective. The three controlled studies are expected to settle the question. Cooper's study, which showed poor results, is cogently attacked. Lang and Lazovik's experiment on college students with fears of snakes comes out all right but it is a far cry from the clinic. Lazarus compared group desensitization with other group techniques. He found the desensitization techniques much more effective, and when he then applied them to his control groups they also improved. It is only then that one notices that the same therapist apparently gave all types of treatment, yet no doubts are raised about contamination or relative skill with different techniques. The magic 'two-thirds' is again not discussed when arithmetic shows that of a total of 33 patients, 23 recovered but only 19 maintained this improvement. Indeed the authors claim that 'the evidence which [the control studies] provide is very striking' (p. 256). The only hopeful finding which consistently turns up throughout the literature is that symptom substitution is rare.

There is one howler: it is implied (p. 2) that psychiatrists rarely give drugs but often use E.C.T. in neurotic disorders. For those with long memories it is interesting to see that hysterics are no longer the *criterion* group for extraversion: (p. 28) 'for the purposes of the discussion, therefore, we shall *henceforth treat* them as extraverts...' (my italics). There is some sleight of pen at times: p. 283, '...we have used the terms 'sympathetic' and 'parasympathetic' in a rather inexact shorthand notation to refer to hedonically positive and negative experiences, respectively; the reader familiar with the complexities of autonomic reactions will no doubt be able to translate these blanket statements into more precise language appropriate to each individual case'.

In conclusion the book makes little theoretical advance, offers little practical help except as a bibliography, repeats a great deal that has been said elsewhere with little sparkle in style and shows dedication to and special pleading for a cause that can stand without it. The most interesting point it brings out is that spontaneous remission sets problems that theorists have not taken seriously enough before.

P. R. F. CLARKE

Abnormal Psychology: Understanding Behaviour Disorders. BY J. R. STRANGE. New York: McGraw-Hill. 1965. Pp. vii+488. 54*s*.

This is a beautifully made text-book in the American style intended for psychology students, both 'majors' and 'non-majors'. Although superficially subscribing to the 'psychobiosocial' approach, the author appears fully committed to a psychodynamic view. The contributions of both experimental psychologists and physically oriented psychatrists are thoroughly neglected.

The book begins with an historical introduction which also serves to present potted Freudian theory. The 'contemporary approaches' include the systems of Horney, Fromm, Sullivan, the existentialists (May), Goldstein, Rogers and Dollard and Miller. To this inexpert reviewer, this appears well done, and so does the section on precipitating factors and defence mechanisms which seem to have more flesh and blood reality than many elementary outlines. The next hundred pages are given to descriptions of psychoneuroses, conduct disorders and psychosomatic disorders. The clinical pictures are presented without much systematic theory. No distinction is made between conversion hysteria and hysterical personality; phantom limb pain appears to be regarded as an hysterical conversion symptom. The chapter on psychosomatic disorders reads far less cautiously than the summary which says that only 'tentative cases' have been made out for 'a psychosomatic connexion in migraine, rheumatoid arthritis, Ménière's syndrome and cancer'. (p. 236).

The flavours of different psychotherapeutic systems are conveyed quite well in two chapters: these include some extraordinary techniques. Wolpe gets a mention but the general concept of behaviour therapy is not discussed: 'symptom-chasing' is castigated. Eysenck's critical statistics are exhibited but neither accepted nor rebutted.

The psychoses are dealt with in two chapters. The approach is clearly psychodynamic: E.C.T. seems to be given to manic-depressives in order to suppress symptoms sufficiently to allow the psychotherapist to get to grips with the problem.

'Somatogenic disorders', mental deficiency and eppilesy are treated sketchily. The mental-deficiency section is up to date on the number of chromosomes in mongols but gives no impression of the excellent experimental work on incentives and training that has changed the face of mental subnormality practice. 'Mental retardation' is used mysteriously to mean temporary lowering of intelligence by emotional disturbances. The uncritical claim is made that 'diet corrects the hereditary condition of P.K.U.' (phenylketonuria) (p. 411).

Easily the worst chapter in the book is the one on 'somatic therapies'. Insulin coma treatment which has been debunked for years is claimed to be 'the most effective treatment for non-chronic simple and catatonic schizophrenics' (p. 435). There is no hint of the recent developments in chemotherapy: the only tranquillizers described are reserpine and chlorpromazine, and the existence of the anti-depressents which have vastly reduced the use of E.C.T. in this country over the last five years is not mentioned at all. There is also the warning (p. 423): 'Tranquillizers should never be used with depressed patients....' Psychosurgery gets a brief section and an illustration of a thoroughly moth-eaten brain which had been 'lobotomized' (p. 431). Hydrotherapy is described in some detail and mention is made of Metrazol shock treatment, heat-treatment and massage.

There are other features which may puzzle the British reader. The staffing of British hospitals and the qualifications of clinical psychologists is different from the scheme set out in the introduction. The classifications and terminology often have a foreign flavour: American government statisticians seem cheerfully to accept the diagnosis of involutional psychosis in a substantial number of people under the age of 25 (p. 23).

As an elementary book about the psychodynamic approach to psychiatry this is probably quite good. As a text on abnormal psychology for psychology undergraduates in this country it has serious flaws.

P. R. F. CLARKE

Abnormal Psychology. By EPHRAIM ROSEN and IAN GREGORY. Philadelphia and London: W. B. Saunders. 1965. Pp. vi + 553. Price not stated.

The authors, a psychologist and a psychiatrist, have brought together, and tried to integrate, the theories and findings of psychoanalysis, experimental psychology and psychiatry in order to provide a comprehensive text for undergraduate students of psychology and medicine. The mixture is much the same as usual, a few chapters on the causes of disorders, hereditary, biological, psychological and 'socio-cultural', and a few on the syndromes, psychoneurosis, affective disorders, schizophrenia, sociopathic disorders, etc. Most of the material is familiar, and there is little or nothing which is not readily available elsewhere. The book is lengthy, being over a quarter million words, perhaps too lengthy for most courses in the U.K., but not detailed, critical or documented enough to serve as a reference book. Nevertheless, the authors have done reasonably well what they set out to do, and the result is a good example of its kind. It follows the fashion of trying to present fairly and reconcile several points of view. Is this what students should be taught to do? There are good reasons why the viewpoints of psychoanalysis, experimental psychology and psychiatry are different. Is one-sidedness a fault? It is not better to teach students to sustain a viewpoint and to take a long, hard, and perhaps penetrating look at the questions?

D. RUSSELL DAVIS

The Role of Psychosomatic Disorder in Adult Life. Edited by J. O. WISDOM and HEINZ WOLFF. London: Pergamon Press. 1965. Pp. xiv + 229. 60*s*.

Principles of Treatments of Psychosomatic Disorders. Edited by PHILIP HOPKINS and HEINZ WOLFF. London: Pergamon Press. 1965. Pp. x + 118. 35*s*.

Psychosomatic Disorders in Adolescents and Young Adults. Edited by J. HAMBLING and PHILIP HOPKINS. London: Pergamon Press. 1965. Pp. ix + 244. 60*s*.

These three volumes are the proceedings of conferences held by the Society for Psychosomatic Research in the years 1960, 1961 and 1962. In their respective introductions the editors bewail the unexpected and unfortunate delay in appearance of the volumes, but feel that as the problems that are dealt with are fundamental the contributions do not date.

The authors of the papers cast their nets widely, and contributions may consist for example of pure psychoanalytical psychopathology, somatic medicine, neurophysiology or clinical psychology.

Most of the papers are interesting in themselves and of a much higher critical standard than one usually gets in reports of symposia. On the other hand, their very diffuseness makes it difficult to know exactly what value comes from such publications. Together they are not sufficiently systematic to make, as it were, a text-book of psychosomatic medicine, nor is there more than a very occasional short bibliography. Nevertheless, these books give a livelier picture than is obtained anywhere else of some growing points of research and treatment in psychological medicine.

D. A. POND

Mental Testing in Clinical Practice. By MOYRA WILLIAMS. Oxford: Pergamon Press. 1965. Pp. xix + 177. 21*s*.

To undertake a topic like this in a book this size is a hazardous enterprise. In most sections of her book Dr Williams succeeds in saying clearly a great deal that needs to be said, but in some the reader is left wondering what purpose the condensation serves. The book has chapters on intelligence, personality, speech and language, memory and learning, perception and orientation, and on the future of clinical psychology. Except for the last, each chapter begins with a discussion of definitions, basic theory and factors affecting the concept under discussion, and ends with brief descriptions of a number of tests used. Nowhere is there any evidence of taking wooden nickels and it is precisely in her cautions and criticisms that she exhibits her clinical good sense, realism and respect for her patients. There is a brief introduction giving an historical perspective and discussing concepts basic to psychological testing such as standardization, reliability and validity. The final chapter makes some shrewd observations on why psychologists do not do research and insists that they should.

In considering the value of the separate sections one has to take into account the type of reader for whom they would be useful. The chapters on intelligence and personality would not be useful to trainee clinical psychologists except for the section on testing old people where the book comes to life, and perhaps the part on tests of deterioration. On the other hand, the lay readers, like general practitioners, medical students, trainee psychiatrists, social workers or teachers for whom the book appears to be intended, would probably benefit more from fuller exposition of the basic concepts (particularly the nature of testing in general) and less listing of tests in terms so brief as to convey little to those who have not some acquaintance with them. The later sections are much more successful and have useful meat for trainee clinical psychologists as well. Indeed one could wish that Dr Williams's points about the breakdown of memory processes might be stamped indelibly on the cuffs of every white coat in the land and scratched deep into the hippocampal gyri (bilaterally!) of every clinical psychologist.

The style may put off both types of reader. The 1, 2, 3, style of condensation and clinical pronouncements will be a joy to anyone trained in the medical tradition and offend the psychologist, whereas the occasional use of technicalities like 'standard scores' and liberal use of references which need (and deserve) to be followed up may work in the opposite sense. The selection of tests inevitably leaves gaps but this reviewer feels that the 1960 revision of the Stanford–Binet should be included at the expense of losing the Healey Picture Completion and Pinter–Patterson Scale. The Babcock is perhaps worth rescuing from obscurity but the Hunt–Minnesota and Shipley–Hertford (*sic*) could stay in limbo. Similarly it is odd to put in the Cornell Index, Bell's Adjustment Inventory and the Bernreuter while leaving out the M.P.I., the E.P.I., and Grygier's Dynamic Personality Inventory. Her line on the Rorschach is that it is useful to use impressionistically; on the T.A.T. she appears a shade uncritical. In the chapter on perception and orientation she seems to use some phrases ('visual agnosia' and 'constructional apraxia') rather more widely than is general. I would take issue with her on some minor points in the chapter on memory but the general message is so good and has so often been neglected that I will not tarnish her shining example. There are over 300 references, of which no less than 12 are to the author's own valuable work.

The paper and type-face are pleasing but these qualities are marred by many irritating misprints and a rather insecure soft cover.

This is a good book and my negative criticisms are only put in the hope that Dr Williams may be persuaded to write a further, more technical account for her psychological colleagues who in general neglect the neurological field.

P. R. F. CLARKE

Family and Class Dynamics in Mental Illness. By JEROME K. MYERS. New York: Wiley. 1959. Science Editions, 1964. Pp. xi + 295. 15*s*.

This book, originally published in 1959 and now attractively reprinted as a paperback, was the second part of a study of the interrelation of social class and mental illness in New Haven (Connecticut), the first part having been Hollingshead and Redlich, *Social Class and Mental Illness* (1958). Because of the paucity of material on the sociology and social psychology of the British family, an American paperback such as this may have an appeal which outweighs its intrinsic merits. It is not too long; it has a plausible scientific air about it even while it relegates all statistics to an unpublished limbo; and it deals in detail with aspects of life which are of great interest. Indeed, Josephine Klein has made much use of Myers and Roberts in her two-volume work on socio-psychological patterns of class and family in England, (*Samples from English Cultures*, London: Routlege and Kegan Paul, 1965). In her discussion of middle-class child-rearing practices she says this of the relevant part of Myers and Roberts: 'Because it is the most detailed analysis available and because the family pattern and its consequences do seem to reappear in English accounts, we shall use their findings in spite of the oddity of the population, in spite of the fact that it is an American study, and in spite of the fact that the social position of these families is somewhat higher than that of the English population on which there is equivalent information.'

The reader should bear in mind the caveats implied in Miss Klein's remarks as well as those set out by Myers and Roberts in their concluding chapter. The study, based upon interviews with mental patients and their families and detailed analyses of their reconstructed life-histories, was limited to a sample of fifty cases drawn from two social classes (classes III and V as determined by Hollingshead's Index of Social Position—a method which does not yield classes comparable with those of the Registrar General). It deals with treated schizophrenics and psychoneurotics only; there was no control group of 'normal' or non-psychiatric cases. The method was retrospective and depended upon recall by informants and patients of the details of the patients' earlier life-histories; incidentally, it may be noted that very few data on pre-school age patterns were thrown up.

There is no discussion of the clinical conditions and the prognoses of the patients at the time of the research, nor is there any detailed consideration of the consequences of the mental illness for the patient or his family. It is as if the interviews with schizophrenics, neurotics, and their families were accepted at face value, and the resulting life-histories accepted as revealing typical patterns of social behaviour, which are then said to be related to mental illness because they are found in schizophrenics and neurotics. Rationalizations of the mentally ill become the theoretical framework for aetiological speculation: thus we are told that class III schizophrenics were shy and somewhat withdrawn as schoolchildren, that they concentrated on their studies, achieved higher levels of educational attainment than their parents and siblings, and that subsequently they were extremely frustrated when their higher educational and occupational ambitions were not fulfilled. This frustration is asserted to have been more severe than that of class III neurotics, although there is no evidence of any attempt having been made scientifically to measure frustration. Schizophrenia is also considered to be a more severe illness than neurosis, again without any evidence, and we find the authors stating: 'Our finding that the greater the frustration in the patients' developmental history, the more severe the psychiatric disturbance supports the theoretical position that extreme frustration is related to the development of schizophrenia.'

It appears that Myers and Roberts envisage frustration as an antecedent factor; they do not appear to have considered whether the shyness of these schizophrenics at school was not a sign of the disease and their occupational frustration a consequence. They have confused the behavioural consequences, and unfortunately they chose to study the former by a method appropriate to elucidating the latter.

G. F. REHIN

Criminal on the Road: A Study of Serious Motoring Offences and Those who Commit Them. By T. C. WILLETT. London: Tavistock Publications. 1964. Pp. xv + 343. 42*s*.

It is a popular preconception that motoring offences differ in some way from other criminal offences: that, unlike larceny, for example, they are 'there-but-for-the-grace-of-God' phenomena euphemistically described as 'accidents'. In *Criminal on the Road* Dr Willett examines the validity of this view and reports on a sociological survey of six types of serious motoring offence committed over a selected period in a typical Police District of the Home Counties. The investigation is divided into two parts: a documentary study of 653 cases extracted from police records; and an interview study in which 43 traffic offenders were personally interviewed by the author.

The documentary study effectively destroys the popular image of the errant motorist as the otherwise law-abiding, white-collared owner of a private car. In fact, of those cases surveyed 62 per cent came from manual occupations, a disproportionate number were motor-cyclists, a fifth committed a further similar offence, and, most important of all, the percentage with a criminal record of non-motoring offences was substantially higher at 23 per cent than even the most pessimistic figure of 1 per cent expected in the general population.

The interview study supports these findings but adds little to them. The psychologist coming to the book will, in fact, feel disappointed that, having gained the co-operation of such a group of people, the author did not take the opportunity of assessing personality factors objectively. This might have proved extremely valuable in view of some recent hypotheses linking criminality and personality, and of evidence that traffic offenders show a high incidence of personality abnormality. This criticism does not, however, detract from the value of what was intended to be primarily a sociological survey of a form of human misdemeanour that is surrounded, perhaps more than most, by irrational thinking and legal inconsistency.

The book deserves attention not only for the facts that its author has unearthed, but also for his excellent analysis of the legal and social aspects of motoring and of the historical developments that have led to the current attitude, on the part of public and police alike, towards motoring offences. As Dr Willett points out, the public image of the motoring offender is maintained by the large number of trivial offences brought before the courts. These shelter a hard core of serious offenders who are legitimate subjects for study by criminologists and also, it is hoped, by those psychologists interested in applying personality theory to the problems of criminality.

G. S. CLARIDGE

Young Men in Detention Centres. By ANNE B. DUNLOP and SARAH MCCABE. London: Routledge and Kegan Paul. 1965. Pp. xii + 180. 28*s*.

The unfashionable doctrine of deterrence was the guiding principle of the most recent and controversial development among penal methods for young offenders—the detention centre. The result, starkly portrayed in this book, is a régime which combines the pointless rituals of military training—parade-ground drill, inspections, endless changes of clothing and the rest—with the corrosive subculture of a closed prison. Characteristically, the Prison Commissioners neglected to make provision for systematic study of this new penal device when it was introduced; it is to the credit of their successors that they gave facilities for a belated attempt to 'evaluate the work of detention centres...and to see what effect this form of punishment had upon those who were subjected to it'.

Despite this challenging brief, Dunlop and McCabe chose to concentrate their attention upon the social attitudes of inmates and custodial staff. They are impatient of orthodox attitude scales, preferring to rely upon 'methodically prepared interviews on lines used in modern social casework'. To judge by the interview schedules set out in the appendices, the leading questions were direct and simple. Except for a few illustrative examples, responses are not presented *verbatim*, but classified by the interviewer into a small number of categories. It is thus impossible for the reader to check for himself the extent to which the interpretations offered are justified by the material obtained. Some of these are of the speculative 'perhaps...but then again, perhaps not' variety which will be familiar to connoisseurs of casework literature: 'The self-conscious use of a recently learned jargon gave evidence of the willingness with which some men identified themselves with the more criminal elements at each centre, although it is possible that this may be due to a naiveté that responded to the piquancy of a new language.' Occasionally the authors seem to contradict themselves: thus the section dealing with 'Intelligence shown during training' begins: 'The chance of a man adjusting to life in the detention centre...may be greater or less according to his endowment of intelligence and the level at which he manages to function'—apparently implying some correlation between these variables—but concludes: 'Dullness does not appear necessarily to be a handicap in training nor high intelligence an advantage but an air of good sense and responsibility clearly inspire confidence.'

The reader closes this book with mixed feelings. One cannot but be struck by the sterility of an evaluative study which contains no hard facts about the correctional effectiveness of the régime, yields no objective criteria for distinguishing those for whom it is suitable from those for whom it is not, and indeed makes no real attempt to relate attitudes during training to the probability of subsequent recidivism. These are surely basic questions that must be asked of any new penal method, and they demand much more sophisticated methods of investigation than were used here. But for all its methodological shortcomings, it is an engaging book; the authors have caught something of the atmosphere of the detention centre and set it down in vivid prose. It is well worth reading for this alone.

GORDON TRASLER

Roles: An Introduction to the Study of Social Relations. By MICHAEL BANTON. London: Tavistock. 1965. Pp. x + 224. 21*s*.

The subject of 'roles' is common ground for sociologists and social psychologists, the latter being indebted to the former for this conceptual tool. Most text-books of social psychology make a gracious bow in the direction of roles, but rarely go beyond a few generalities coupled with references to such limited psychological studies as are available. This is regrettable, because the concept is useful in the analysis of large areas of social behaviour; and even the psychologist primarily interested in the behaviour of individuals needs some knowledge of prescribed role behaviour as a baseline. Prof. Banton goes so far as to claim that the concept can help 'in the analysis of any kind of conscious behaviour'. The reviewer's first reaction was to register a protest in the margin; on second thoughts, it seemed that perhaps the potentialities of the concept have as yet been insufficiently exploited: the role of the experimenter's subject bears looking into more systematically, for instance. Nevertheless, to say that all conscious human behaviour has to do with roles is saying too much, or else the concept becomes redundant. It follows that we need a definition, preferably operational, in order to distinguish role-behaviour from other types. Like his predecessors, the author fails to provide this, and perhaps it just cannot be done. On the other hand, he proposes a new approach to the classification of roles in terms of their degree of 'differentiation' as opposed to content, as usual hitherto. This results in a scale, provisionally assumed to be undimensional, from 'basic' roles (sex, age) via 'general' roles (e.g. occupational) to 'independent' ones (leisure). Closer examination suggests that the scale is unlikely to be strictly undimensional, since the notion of differentiation is used in at least three distinct ways: (1) degree of independence of a role from the social structure; (2) the range of social situations over which a given role is relevant; (3) the amount of freedom the incumbent of a particular role has either in performing this particular role, or in adopting other ones. Although clearly aware of these complexities, Prof. Banton tends to skate over them somewhat too lightly. Apart from these reservations, a new departure is to be welcomed, and one agrees with the author that the real test is whether it will open up new problems for research.

Because of its importance, the reviewer has devoted a large part of his allocated space to one single chapter out of nine. In some of the others the theoretical formulation is applied in such contexts as the consequences of economic development (from a few basic roles to many independent ones) or the signs whereby different types of roles are identified. One of the most rewarding chapters deals with studies of roles and role-conflicts in different professions. The discussion is wide-ranging and always lively, including such topics as prestige and social class. Illustrations are drawn from societies far apart in time and space, with contemporary Britain fortunately figuring prominently.

Prof. Banton is an acute observer of the subtleties of social relations, and describes them with a perceptiveness which social psychologists might well emulate, though they would wish to go on and formulate more specific hypothesis capable of being tested; one of these gems for instance, dropped in passing, concerns signal behaviour, whereby people convey their intentions to others without being aware of it.

The book as a whole makes enjoyable reading and offers numerous valuable insights, not least about the contrasting approaches of sociologists and psychologists to essentially the same raw material, human behaviour.

GUSTAV JAHODA

New Directions in Stuttering: Theory and Practice. Edited by DOMINICK A. BARBARA. Springfield, Ill.: Charles C. Thomas. 1965. Pp. 188. $7.25.

This monograph describes a fair range of approaches to the problem of stuttering in the United States.

The first chapter by Smiley Blanton outlines a psychoanalytic approach. There are other psychoanalytic approaches which are at variance with some of the views expressed. Blanton's work is also at the opposite pole to recent work in Britain, where the pendulum is swinging back towards a more conscious emphasis on speech.

Several of the chapters attempt to use a more scientific language in organizing the concepts the writers wish to put forward, culminating in a scheme by Shames and Sherrick to fit stuttering into the framework of operant conditioning. Within this framework an essential assumption is that stuttering has its origins in the ordinary non-fluencies of childhood, which, by a process of positive and negative reinforcement on complex, multiple schedules, may develop into conditioned repetitions in speech. While this may allow for testable hypotheses to be formulated, the basic assumption is still not proven and itself requires more exploration. Such exploration is attempted by Bloodstein, Alper and Zisk, whose chapter on stuttering developing from normal disfluency presents some evidence to support such a relationship. However, the authors themselves acknowledge the data to be insufficient to warrant any conclusive statements.

Cooper's chapter outlines a therapeutic approach called IPC (Inter-personal Communication). He stresses the poor inter-personal relationships of stutterers and says the relationship between the stutterer and the therapist is the most important factor in therapeutic success. He claims that IPC defines the therapeutic relationship more precisely and aims at bringing therapy into more quantifiable proportions. It is still difficult to see how the technique lends itself to experimental verification any better than other forms of psychotherapy.

Martin, Ward and Earle Johnson's chapter on 'The self as a central concept in speech therapy for the person who stutters' will interest those concerned with the work of Lewin, Sullivan and Rogers.

Millburn Clark and Pemberton Murray outline a technique to study the self-concepts of stutterers by means of autobiographical details and tests, which seems rather cumbersome.

Perkins looks for common denominators in stuttering and formulates some rather untestable hypotheses concerning conflict in relation to stuttering.

The final chapter, by the editor, 'An experiment into the team approach to group psychotherapy with stutterers', advocates both speech therapy and psychotherapy. The symptomatic approach favoured is that of Bluemel. Group psychoanalysis combining stutterers and non-stutterers within the group is carried out conjointly. An investigation of ten stutterers is presented.

The book is of interest and contains ideas into which further research is warranted. Even if one feels disappointed that, with the volume of work going on in the United States on stuttering, the baffling problems of both aetiology and treatment remain so far from solved, it is more depressing to contemplate the dearth of research going on in this country by comparison.

BERENICE KRIKLER

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp*. 225–238
Printed in Great Briatin

THE ALTERNATION PROCESS IN BINOCULAR RIVALRY

By W. J. M. LEVELT

Institute for Perception RVO–TNO, Soesterberg, The Netherlands

Alternation frequency in binocular rivalry and relative dominance of stimuli in the right and left eyes are described in terms of an alternation model. The model is based on the assumption that the mean duration of the dominance of the stimulus in one eye is independent of the strength of this stimulus; the duration is assumed to be dependent only upon the strength of the stimulus in the contralateral eye. A provisional definition of stimulus strength is given. Evidence for assumptions and model is presented by a review of experimental literature on dominance and alternation in binocular rivalry, and by a number of experiments. Normal binocular fusion is considered.

In a previous paper (Levelt, 1965*a*) the author described some experiments which explored binocular brightness impression as a function of the respective luminances of the two monocular fields, and of an extra contour in one of these fields. Two mechanisms of binocular interaction appeared from this study. First, binocular brightness was constant as long as a sum of weighted left and right test field luminances was constant: $w_l E_l + w_r E_r = C$ where E_l and E_r denote left and right test field luminances (energy) respectively. It was shown that the sum of the weighting coefficients $w_l + w_r$ was constant for all of the situations studied. Stated otherwise, if the weighting coefficient for the field in one eye is increased, the weighting coefficient for the corresponding field in the other eye is decreased in the same measure. This was called the *law of complementary shares*. Secondly, the weighting coefficient for an eye can be increased by placing a contour in the monocular test field of the eye. The contribution of the other test field is reduced correspondingly. It was shown that the extreme case could be approximated in which w for one eye tends to zero, as is the case if the other eye is fixating at or near a contour. This was called the *contour mechanism*; introduction of a monocular contour increases the weighting coefficient for that eye. In classical rivalry situations, e.g. a binocular crossing of bars, these mechanisms are incompatible. A contour is presented to both eyes, therefore both weighting coefficients tend to increase according to the contour mechanism. However, according to the law of complementary shares this is not possible, because the sum of the weighting coefficients is constant. The apparent resolution of this perceptual conflict is that the law of complementary shares is saved by an alternating process. Each eye, in its turn, has a proportional contribution of unity. This process of alternation can be influenced by a number of stimulus variables which seem to be related to the contour mechanism. It is reasonable to expect that stimulus conditions determining dominance of a monocular field in binocular brightness averaging also determine dominance in the alternation process.

The literature on alternation in binocular rivalry shows two major shortcomings. First, although the effect of a number of variables on the alternation process has been studied, there has been little attempt to propose underlying mechanisms that might explain how different variables can produce very similar effects. For instance, blurring the monocular stimuli and decreasing their luminance have the same effect on alternation; the alternation rate is reduced. Are they instances of the same underlying process? A second shortcoming is the following. Since Breese (1899, 1909) two features of the process have generally been studied. The first is the alternation frequency. The second is the dominance aspect of the process, i.e. the part of the total viewing time in which the right and the left stimuli are visible. Nobody has ever proposed a model which unifies both aspects of the same alternation process.

In this paper suggestions are made to fill in these gaps. A concept of stimulus strength is introduced to unify the effective variables, and a model is used to describe the alternation process of binocular rivalry in both its frequency and dominance aspects. The combination of stimulus strength, as defined, and the model lead to a number of testable deductions.

Stimulus strength

It is assumed that the features of the alternation process are determined by a uni-dimensional variable, which will be called *stimulus strength* (λ), and which will, where possible, be defined up to order relations. As noted, we found in our previous study that the weighting coefficient of a monocular area for the binocular brightness impression increased when the mean distance between fixation point and a monocular contour near that area decreased; i.e. the effect of a contour declines with distance, or, alternatively, the weight is a function of the amount of contour per area. This leads to the assumption that stimulus strength in binocular rivalry is a monotonic increasing function of the *amount of contour per area*. Furthermore, our previous data suggested that the 'radius of action' of a contour diminishes when its contrast is reduced. Hence we assume that stimulus strength is directly related to the effectiveness of a contour: '*contour strength*'. The intensity of a stimulus is conventionally defined as the ratio of stimulus to threshold-value. The equivalent definition of contour strength is the ratio of contrast to the difference threshold.

Starting from such a concept of contour strength, a large number of stimulus variables can be conceived of as special cases of contour strength variation. There are, then, two obvious ways of increasing contour strength: first by increasing *contrast*, and secondly by reducing the contrast difference threshold. This threshold depends on a large number of conditions.

A few of them are pertinent. Threshold contrast decreases with increasing *luminance* of the test object up to some value (about 20 trol.), but for high luminances a slight increase is often found probably due to saturation effects. Thresholds rapidly increase with the degree of eccentricity of the stimulus. Difference thresholds are smaller for large *sizes* of test object. The difference threshold increases if *blur* is introduced in the contours (see, for example, Ogle, 1961). And when the contrast is a combined colour and luminance contrast, our measure for contour strength is equivalent to the distance of the two colours in the colour space. This distance depends, in a complex way, on the *state of colour adaptation* of the eye.

For all experiments found in the literature, the conditions are scalable along such a dimension of ordered stimulus strength. And although the alternation model, which will now be developed will only be checked up to order relations, the model lends itself equally well to the use of metric relations.

The alternation model

The merits of the model to be developed are independent of the definition of stimulus strength. It comprises only the relation between dominance and frequency aspects of the alternation process. For instance, stimulus variation A appears to affect alternation frequency if applied to both monocular stimuli: the model then predicts particular changes in dominance and rate of the alternation process, if the same variation A is applied to only one of the monocular stimuli.

A contour presented to one eye is absolutely dominant as long as the other eye is presented with a homogeneous field. There are no 'saturation effects' in this situation, in the sense that the contour fades at times in favour of the homogeneous field. Sometimes it looks as if the latter phenomenon does occur, but this is due to Troxler's effect, and is not a result of binocular interaction. It is, therefore, a case of 'spurious rivalry' (Levelt, 1965*b*). Variations in contrast or luminance do not alter the general fact that the perception from one eye continues without interruption as long as no rival contour is presented to the other eye. This fact may have important implications for a model of binocular alternation. Thus, it is plausible that the mean period during which one of the stimuli is visible during binocular rivalry should not be a function of the strength of the same stimulus, but only of the strength of the contralateral stimulus.

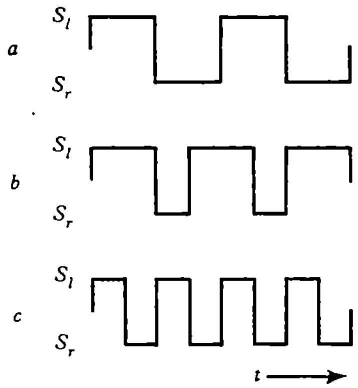


Fig. 1. The alternation model. Effects of variations in stimulus strength over time, t: (*a*) $\lambda_l = \lambda_r$ giving $\bar{t}_r = \bar{t}_l$; (*b*) increase of λ_l only leading to reduction of $\bar{t}_r$ and increase of frequency; (*c*) increase of both λ_l and λ_r leading to reduction of both $\bar{t}_r$ and $\bar{t}_l$ and increase of frequency.

The following symbols are used. $\bar{t}_r$ denotes the mean time during which the right stimulus is uninterruptedly perceived, and likewise $\bar{t}_l$ denotes the mean period length for the left stimulus. These symbols are used only in relation to the two-choice rivalry task, hence $\bar{t}_r + \bar{t}_l = \bar{T}$, the mean duration of the complete cycle. S_r and S_l are right and left stimuli respectively, and λ_r and λ_l are right and left stimulus strengths. The predominance of a stimulus is defined as the percentage of the total viewing time during which this stimulus is perceived. Hence, in symbols, $\bar{t}_r = f(\lambda_l)$ and $\bar{t}_l = f(\lambda_r)$, with $\bar{t}$ a monotonic decreasing function of λ. Before deducing some propositions, from this assumption, the idea may be stated still otherwise. If S_r is perceived the conditional probability that during some time interval t, S_l becomes visible, depends only on λ_l, not on λ_r, and inversely for a change from S_l to S_r.

The precise characteristics of the distributions of t or the function $f(\lambda)$ are discussed later. Even with $f(\lambda)$ unspecified it is possible to deduce four testable propositions on dominance and rate of alternation (cf. Fig. 1).

(I) *Increase of the stimulus strength in one eye will increase the predominance of this stimulus.* Deduction of the proposition for S_r: the predominance of S_r is $100\,\bar{t}_r/\bar{T} = 100\,\bar{t}_r/(\bar{t}_r+\bar{t}_l)$ but $\bar{t}_l = f(\lambda_r)$, so an increase of λ_r leads to a reduction of the denominator, and this means an increase of the predominance of S_r.

(II) *Increase of the stimulus strength in one eye will not affect $\bar{t}$ for the same eye.* This is a negative but important statement: $\bar{t}_r$ can only be affected by λ_l, not by λ_r.

(III) *Increase of the stimulus strength in one eye will increase the alternation frequency.* Increase of λ_r reduces $\bar{t}_l$, and leaves $\bar{t}_r$ unaffected, hence $\bar{T} = \bar{t}_r+\bar{t}_l$ is reduced, and this is equivalent to an increase in rate of alternation.

(IV) *Increase of the stimulus strength of both eyes will increase the alternation frequency.* Increase of λ_l and λ_r leads to reduction of both $\bar{t}_l$ and $\bar{t}_r$, and hence of $\bar{T}$.

REPORTED DATA ON ALTERNATION RATE AND DOMINANCE

Most of the reported work of binocular rivalry is concerned with alternation rate and dominance. A vast majority of the experiments are of the dominance type: the stimulus conditions are varied for one eye, and the effect on the predominance of that stimulus is measured. A smaller part of the literature is concerned with simultaneous changes of the conditions in both eyes: these studies normally give data on alternation rate. A minority of the studies give data on rate of alternation for the situation in which the stimulus presented to one eye is varied. A somewhat different classification is used below, because the model relates both rate and dominance aspects of the alternation process; the number of the proposition (I to IV), relevant to the experiment in question is added in brackets.

Contrast experiments

In most experiments contour strength has been varied, sometimes by varying contrast, sometimes by varying the difference threshold. Roelofs & Zeeman (1919) were the first to stress the importance of contrast in binocular rivalry. Using orthogonal grids as stimuli, they showed that the dominance of an eye was reduced, if the contrasts in its grid were reduced, but irrespective of the absolute luminance level of this grid (I). Gellhorn and his collaborators found the same for coloured stimuli (I) (Gellhorn, 1924; Gellhorn & Kuckenberg, 1924; Gellhorn & Schöppe, 1924, 1925*a*, *b*). Mull, Armstrong & Telfer (1956) used stimuli similar to Gellhorn's—a red and blue square on a neutral background presented to corresponding areas of the two eyes. They varied the contrast of the stimuli by using a grey or a black background. This did not notably affect the alternation rate: however, it is not possible to reconstruct what change in contrast took place by changing the background; probably only the contrast reversed, but the sign of the contrast is irrelevant for coloured stimuli, as Gellhorn showed (III). Alexander (1951) and Alexander & Bricker (1952) found an increase of alternation rate if the contrast of the stimuli (grid patterns) was increased in both eyes (IV). Kakizaki (1960), using similar orthogonal grids as stimuli, found an increase of predominance with contrast for one eye (I); the rate of alternation increased with increasing contrast of the pattern (III). Bright bars on a black background were used as stimuli by Kaplan & Metlay (1964). Increasing the luminance of the bar in one eye led to dominance approaching 100 % (I), nevertheless alternation occurred even when the two fields were very different in brightness. These authors also noticed an increase in rate of alternation (III), and an increase in rate with increased illumination of both bars (IV), but increased luminance above 4 log trol. slightly decreased the rate. This accords fairly well with the saturation hypothesis for a decrease in difference threshold at high luminances.

Amount of contour per area. Allers (1935) found an increase of predominancy of an eye, if the amount of contour per area for that eye was increased (a pattern of straight bars was replaced by a pattern of waved bars) (I). Similarly, Alexander (1951) found a decrease of rate, if gaps were introduced in the bars of the two orthogonal grids that he used (IV). A similar example is to be found in Gellhorn's work (*op. cit.*). Studying the effect of colour adaptation in rivalry, he presented two differently coloured squares on corresponding points of the two eyes. He now introduced an additional stimulus of some other colour near the square in one eye, and found an increase of dominance for that square. To his surprise this appeared to be independent of the colour of the additional stimulus. In our opinion the matter is simply one of introducing extra contours (I). A direct test of the radius of action of a contour has only been made by Kaufman (1963). He presented a horizontal bar to the left eye, and two vertical bars to the right eye, so that the binocular impression was an intersection of a horizontal and two vertical lines. The observer passed a key whenever the part of the horizontal bar between the vertical bars disappeared. The amount of suppression appeared to be a function of the angular separation θ of the two vertical bars (I). Complete suppression of the segment hardly occurred beyond $\theta = 2°$. The suppression stabilized at about 50 %, for angular separations within 14 min of arc. Kaufman also found an increase in rate with a decrease of θ (III).

Luminance. There is not much unanimity on the role of luminance *per se* in rivalry. The results of experiments in which a constant (and finite) contrast was used at different luminance levels are rather divergent. This is not surprising, because the difference threshold is a curvilinear function of luminance. For low luminance levels a small luminance change corresponds to a large change in threshold, and hence to a large change in contour strength. For higher luminances the threshold is practically constant, and at some level a slight increase in threshold is often measured. The divergence of the results may be due to the differences in chosen luminance levels. Breese (1899, 1909) found an increase of alternation frequency by increasing the luminance of the two grid patterns used (IV), and a shift in dominance if the luminance was increased for one eye only (I). Roelofs & Zeeman (1919) did not find effects of luminance level variations, whereas Mull *et al.* (1956) did find such effects on alternation rate (IV). But none of these authors give absolute values for the luminances used, and one of our experiments was therefore concerned with this point.

Eccentricity of the test target changes the difference threshold. It was again Breese (1909) who presented the first data on this variable. He found that alternation rate decreased towards the periphery of vision (IV); the same tendency appeared in a study of Wilde (1938) who used two crossing bars as stimuli (IV). The other results of Wilde's rivalry experiments appear not to be due to binocular interaction but to spontaneous fading of an image in the periphery of vision (Troxler's effect cf. Levelt 1965*b*).

Variations in size of the test target were only studied by Breese (1909). For larger grids, the rate was increased (IV). The main effect was in the region of small sizes, but, unfortunately, visual angles were not given.

Adaptation variables have been studied by Gellhorn (*op. cit.*). For his coloured stimuli, he found a decrease of dominance in rivalry for a stimulus, if the eye was preadapted to the colour of that stimulus. The colour contrast is reduced in this way (I).

Blur effects were again initially studied by Breese (1909). In one of his experiments he blurred the contours of the grids in both eyes by out-of-focus projection. In this way not only is blurring of contours introduced, but also reduction of contrast. Hence it is not certain whether the reduction of alternation rate that he found (IV) should be mainly attributed to the effect of blur on the difference threshold; one of our experiments was therefore on blur and rivalry, with contrast kept constant.

The aim of this excursion into the literature was to look for confirmation of the proposed alternation model and for counter-evidence. The booty of the expedition can now be summarized.

1. Proposition I has often been the objective of experimentation and has generally been confirmed, but a simple check of this proposition has more to say about the usefulness of the concept of stimulus strength than about the internal consistency of the model. A check on consistency is only possible if the effects of same stimulus variable are also tested in relation to the other propositions. Examples were given in the work of Breese, who found the predominance-effect (I) and the rate-effect (IV) when monocular and binocular increases in stimulus luminance were introduced. The same effect was found by Kaplan & Metlay (1964); moreover these authors found rate to increase when the luminance of only one stimulus was increased (III). Kakizaki (1960) found effects I and III for uniocular increase of contrast, and finally Kaufman's (1963) data show these two effects for the variable contour-per-area. These studies confirm the internal consistency of the model as far as propositions I, III and IV are concerned. But no data at all have been found which bear on proposition II, the most specific of the propositions. On the contrary, the general opinion seems to have been that the mean duration of a dominance period is *increased* by strengthening the stimulus.

2. There are three components in our tentative definition of stimulus strength: stimulus strength increases (i) with increasing contrast, (ii) with decreasing difference threshold, and (iii) with diminishing angular distance between stimulus contour and target in the other eye. Striking counter-arguments against the use of this parameter were not found, but this is little testimony to its adequacy. More data are needed. The three components have therefore been studied for one type of stimulus: first in relation to distance, then to the effects of contrast and of threshold variations. The last two experiments were also designed to study proposition II in relation to propositions I and III.

The only point of clear controversy appears to be the effect of the absolute luminance level of the stimuli. From our definition of contour strength no luminance effect is expected as long as contrast and difference threshold are constant, for luminances beyond about 20 trol. This point was also subjected to experiment.

EXPERIMENTS ON THE ALTERNATION MODEL

Experiment I. Radius of action

The problem. In studies in which the amount of contour per area is varied, the basic question concerns the 'radius of action' of a contour. The present problem is: how large can the angular separation between a left and a right eye contour be in order still to provide moments of complete suppression of one of them in binocular vision? Kaufman's (1963) data concern foveal vision, no data have been found on peripheral vision.

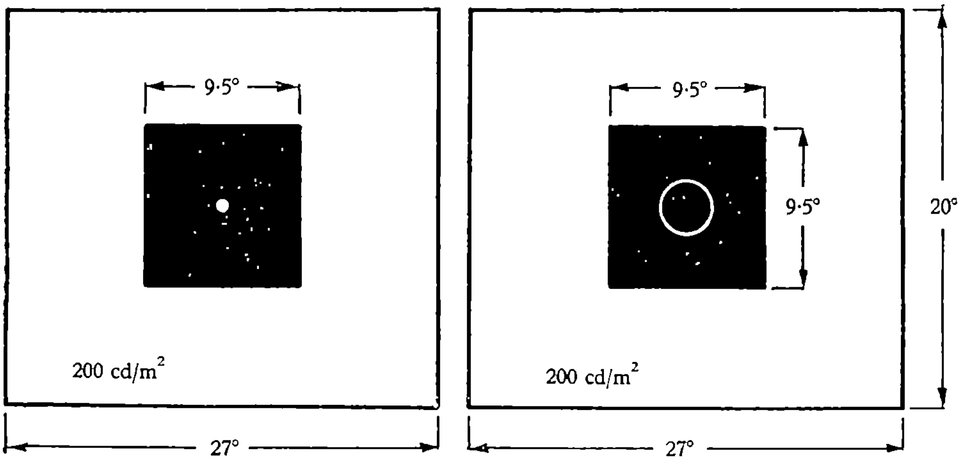


Fig. 2. Stimuli used in Expt I. The point stimulus on the left had a diameter of 20′ (luminance: 200 cd/m²). The circle on the right had a variable diameter and a width of 6′ (luminance: 200 cd/m²).

Method

Apparatus and stimuli

The apparatus has been described in detail elsewhere (Levelt, 1965*a*). The observer looks into a kind of stereoscope through 1 mm artificial pupils. The luminances of the left and right monocular fields can be controlled. The stimuli used in the present experiment are given in Fig. 2. The black squares were fused binocularly. In this way the white (left) test point is prevented from drifting within the white (right) circle. The ring and the test point had a constant luminance of 200 cd/m². We used a ring as 'suppressing' stimulus in order to have a constant distance in all directions between left and right contours.

Procedure

The total disappearance time of the left test spot during a 1 min period of observation was measured as a function of the inner diameter of the surrounding circle. The diameter was varied in 0·5° steps from 4° to 1°. Furthermore, the condition of 'no circle' was tested, for comparison with Troxler's effect. Both central fixation, i.e. fixation of the test spot, and peripheral fixation were tested: in the latter case an extra fixation point was introduced for both eyes, 3° above the test spot. Measurements started with the Troxler condition, followed by a descending series of circle diameters. Two subjects took part in this experiment.

Results

The results are shown in Fig. 3. The radius of action is much larger for peripheral than for foveal vision. For a 3·5° circle, it is distinctly above the Troxler level in peripheral vision, whereas in foveal vision complete suppression only occurs for 1° and 1·5° circles. For all other diameters complete suppression does not occur (although nearly complete suppressions often occur). If we take experimental differences into account, our results for foveal vision are in fair agreement with Kaufman's, as to the maximum extent of complete suppression. In terms of our model, proposition I is confirmed with peripherality and radius of action as components of stimulus strength.

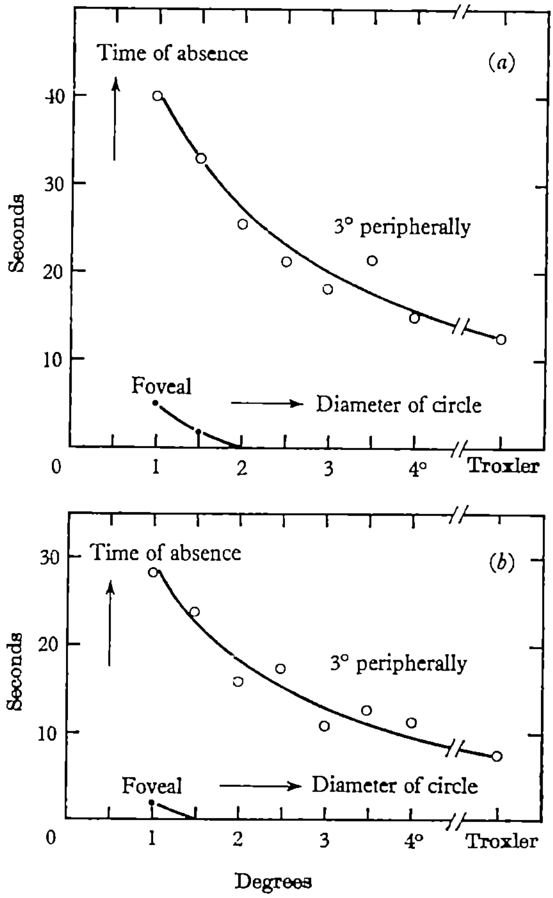


Fig. 3. Suppression of left test-spot as a function of the diameter of the right circle for central and peripheral vision for two observers (*a*) W.L. (*b*) A.v.d.B.

Experiment II. Blur

This experiment and Expt. III were designed to test the relation between propositions I, II, and III of the alternation model.

The problem. Blurring the stimulus in the left field should reduce its contour strength, resulting in a reduction of the predominance of the left stimulus (I), and reduction of the alternation rate (III). However, the mean dominance time $\bar{t}_l$ of this stimulus should not be affected (II).

Method

Stimuli

The stimuli are shown in Fig. 4. The blur-condition B was produced by placing an extra lens in the left light pathway of the apparatus. The result was the luminance distribution shown in Fig. 5. The luminance distribution under the non-blurred condition A is also shown for comparison.

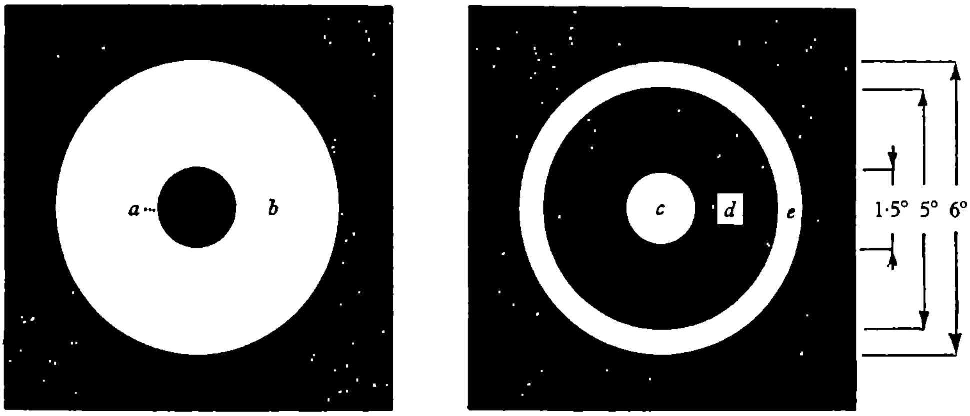


Fig. 4. Stimuli of Expts II and III. The luminances in Expt II were: for Expt II: $a = 40$ cd/m^2, $b = 400$ cd/m^2, $d = 23$ cd/m^2, c and $e = 400$ cd/m^2; and in Expt III: condition A: $a = 85$ cd/m^2, $b = 710$ cd/m^2; condition B: $a = 85$ cd/m^2, $b = 100$ cd/m^2; condition C: $a = 12$ cd/m^2, $b = 100$ cd/m^2; all conditions: $c = 100$ cd/m^2, $d = 5{\cdot}75$ cd/m^2, and $e = 100$ cd/m^2.

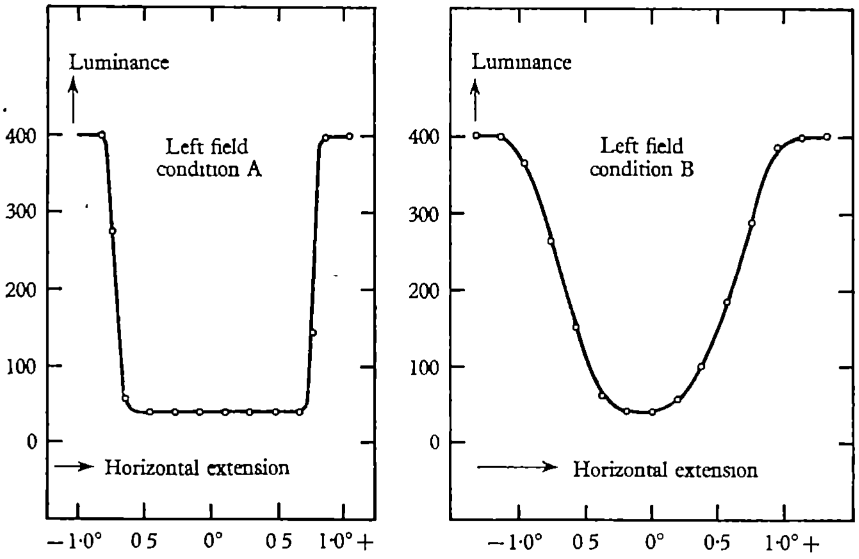


Fig. 5. Luminance density of the dark spot in the left field of Fig. 4 for the non-blur and the blur conditions, Expt II.

Moreover, the total amount of light from an area somewhat larger than a (in Fig. 4, but including it, was measured by means of a photomultiplier. No difference was found between conditions A and B. Hence, under the present conditions, the introduction of blur did not affect contrast or total amount of light of the left stimulus.

Procedure

The subject looked through artificial pupils, diameter 1 mm, and fixated the centre of a and c in Fig. 4. He kept the key depressed for as long as the left stimulus was dominant. Ten subjects took part in the experiment. They observed each condition twice in random order. Each inspection period lasted 1 min.

Results

Two analyses of variance were applied to the results, one to the dominance scores, and another to the frequency scores. Of the dominance scores, the mean predominance for the ten subjects in the unblurred conditions was 50·9 % (or 30·55 sec/min). For the blurred condition, it was 26·5 % (15·87 sec). The difference is significant ($P < 0{\cdot}001$) and confirms proposition I. The mean frequency scores for the A and B conditions were 12·20 and 6·40/min respectively. This difference was significant ($P < 0{\cdot}005$) and confirms proposition III.

As to the proposition II, the mean $\bar{t}_l$ was 2·50 sec in the A condition, and 2·48 sec in the B condition (S.E. = 0·5). No analysis of variance is necessary to appreciate that $\bar{t}_l$ evidently was not affected with the stimulus strength in the right field constant.

Experiment III. Contrast

The problem. This last experiment serves a number of purposes. First, the effect of contrast variation and luminance variation on rivalry were to be compared. The prediction was that, for a sufficient level of luminance, luminance variation would have a negligible effect on rivalry, but that there would be a large effect of variation in contrast. Secondly, separate analyses of predominance and frequency aspects of the alternation would give information about propositions I and III. And finally, calculation of $\bar{t}_l$ would provide another test of proposition II.

Method

Stimuli

Except for their luminances, the stimuli were the same as in Expt II. The luminance values are given in the legend to Fig. 4. Three conditions were used. In conditions *A* and *C*, the contrast of the left test spot was 7·35, but the stimuli differ in the level of total luminance. The luminance of the left field under condition A was 7·1 times larger than its luminance in C. In condition B, the contrast was set at 0·18, much less than under the other two conditions. On the other hand, the luminance level of the stimulus lay between the levels of the A and C stimuli. In fact the test spot had the same luminance as under condition A. and the surrounding field had the same luminance as under condition C. In this luminance region the difference threshold is fairly constant. The right field was the same for all conditions.

Procedure

Ten subjects were presented with the three conditions in random order, with one repetition of the whole cycle, giving six 1 min observation periods per subject. The subject was instructed to push the key for as long as the black on white test spot was visible.

Results

Two analyses of variance were carried out, one on the predominance scores and one on the frequency scores. The average predominance of the left field for condition A was 41·9 % (25·13 sec), for B 27·6 % (16·55 sec), and for C 46·6 % (27·93 sec). Scheffé's method was used to compare these means. It revealed that the differences between B and A and between B and C were significant ($P < 0{\cdot}001$). The small difference between A and C was not significant. These results corroborate the findings of Roelofs & Zeeman (1919) that luminance level is immaterial to dominance in binocular rivalry, as long as contrast is preserved. The effect of changes in contrast appears to be important in that the low contrast in B resulted in low predominance for the left field (I).

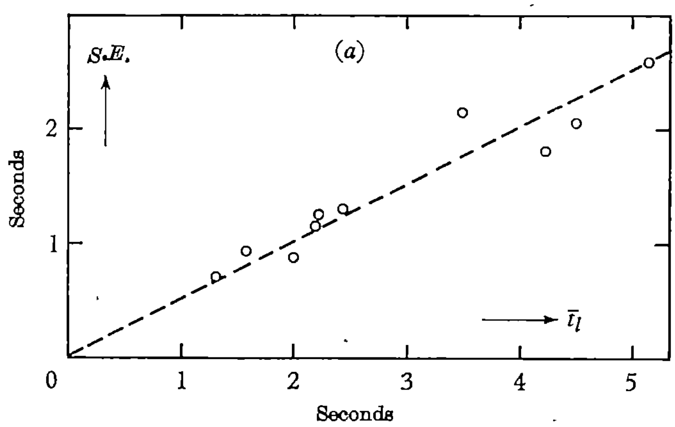


Fig. 6. Relation of mean and S.E. in the blur experiment.

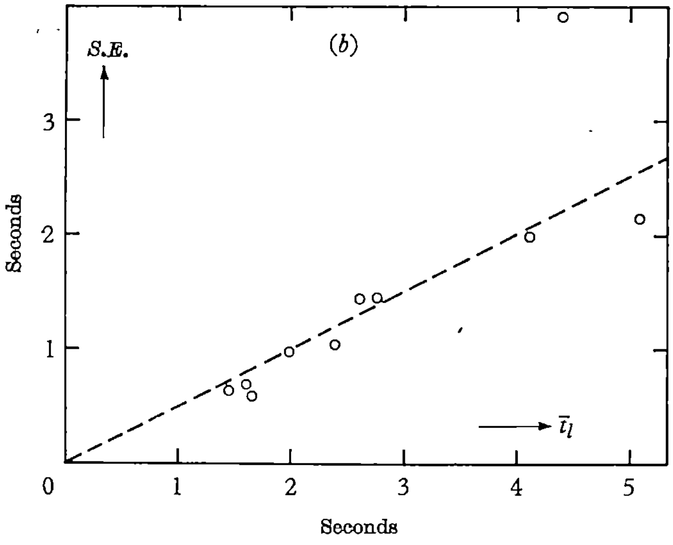


Fig. 7. Relation of mean and S.E. in the contrast experiment.

The analysis of frequency scores again showed significant differences between conditions ($P < 0{\cdot}001$). Mean frequency under condition A was 9·6, under B 6·4, and under C 10·9/min. Again, only the differences between B and A and between B and C were significant ($P < 0{\cdot}001$), in agreement with proposition III.

As to proposition II, for A, B, and C, $\bar{t}_l$ was 2·62, 2·59, and 2·56 sec, respectively, which again confirms the proposition.

CONCLUSIONS

The concept of stimulus strength seems to have merit at least up to order relations. But even at this level there are still gaps in our knowledge. For example, what are the relations between levels of contrast on the one hand and levels of blur on the other hand? Nevertheless, a number of specific predictions could be made, as for instance about the effects of contrast and luminance variation. There is much evidence for an alternation model based on the assumption that the mean duration of the predominance period of one eye is a function only of the stimulus strength in the other eye. Proposition II is supported by Expts II and III. This assumption moreover generated hypotheses about both dominance and rate in the alternation process, none of which had to be rejected.

Fusion and alternation

At the beginning of this paper it was shown that the law of complementary shares conflicts with the contour mechanism if non-corresponding contours are presented to the eyes, but there is no *a priori* reason to expect no conflict if the eyes are presented with corresponding contours, i.e. in normal binocular vision. In the latter case one cannot be aware of an alternation process, because of the law of complementary shares. The impression is exactly the same for all values of w_l and w_r, as long as their sum remains constant.

Authors on binocular rivalry customarily extend the rivalry process to cover the situation of fused images. An early author to do so was Du Tour (1760), and for more than two centuries this point of view has been raised repeatedly. In this century Verhoeff (1935) has assumed that only one of the retinal images is present to consciousness at a time; similar suggestions have been made by Asher (1953) and recently by Hochberg (1964). Panum's fusional area can be understood in this way. Ogle's (1950) measurements on the horizontal extent of Panum's area give values of 6–8′ in the foveal field, that is of the region within which one line is always seen when a pair is presented. These values compare with Kaufman's (1963) data on extent of inhibition: he found a constant value of about 50 % inhibition of the horizontal line segment for $\theta < 14'$ (separation of vertical lines in the contralateral eye). Hence, the extent of complete suppression of a horizontal line in rivalry is also about 7′.

The experiments by Hubel & Wiesel (1959, 1962) have shown that for most cells in the visual cortex of the cat a functional predominance of one of the eyes could be measured. However, absolute dominance appeared to be exceptional. In our earlier experiments (1965*a*) absolute predominance of one eye in binocular brightness averaging also appeared to be exceptional. The proportional contribution of an eye ranged between 0 and 1, but was mostly somewhere in between. Recent experiments by Lansing (1964) showed that the EEG response to a fluctuating left eye stimulus was reduced during periods of perceptual dominance of the right eye stimulus. But this was also no all-or-none effect; different degrees of EEG reduction were measured. It may equally be the case that alternation in binocular vision is not an all-or-none effect, as has been generally assumed. It is suggested that the contributions of the eyes are continuously balanced in the range between 0 and 1, while obeying the law of complementary shares, and that the mean position of the balance is determined by eye dominance and stimulus factors, as discussed in this paper.

REFERENCES

ALEXANDER, L. T. (1951). The influence of figure-ground relationships in binocular rivalry. *J. exp. Psychol.* **41**, 376–81.

ALEXANDER, L. T. & BRICKER, P. D. (1952). Figure ground-contrast and binocular rivalry. *J. exp. Psychol.* **44**, 452–4.

ALLERS, R. (1935). Über einige Unterschiede zwischen dem ein- und dem beidäugigen Sehen und über den Einfluss seelischer Momente auf einfache Leistungen des Gesichtssinnes. *S.B. Akad. Wiss., Wien*, **144** (*math.-naturwiss. Kl.*), 33–79.

ASHER, H. (1953). Suppression theory of binocular vision. *Brit. J. Ophthal.* **37**, 37–49.

BREESE, B. B. (1899). On inhibition. *Psychol. Monogr.* **3**, whole no. 11.

BREESE, B. B. (1909). Binocular rivalry. *Psychol. Rev.* **16**, 410–5.

DU TOUR, E. F. (1760). Discussion d'une question d'optique. *Mém. math. phys. Acad. Roy. Sci. Paris*, **3**, 514–30.

GELLHORN, E. (1924). Quantitative Untersuchungen über den Wettstreit der Sehfelder. III. Zur Kenntnis der psychologischen Ursachen des Wettstreites. *Pflüg. Arch. ges. Physiol.* **206**, 237–49.

GELLHORN, E. & KUCKENBERG, J. (1924). Quantitative Untersuchungen uber den Wettstreit der Sehfelder. I. Der Einfluss der Helligkeit der Farben auf den Wettstreit mit besonderer Berucksichtigung der Kontrastwirkungen. *Pflüg. Arch. ges. Physiol.* **206**, 194–210.

GELLHORN, E. & SCHOPPE, C. (1924). Quantitative Untersuchungen uber den Wettstreit der Sehfelder. II. Die Änderung des Wettstreites durch Umstimmung des Sehorgans. *Pflüg. Arch. ges. Physiol.* **206**, 211–36.

GELLHORN, E. & SCHOPPE, C. (1925*a*). Quantitative Untersuchungen über den Wettstreit der Sehfelder. IV. Über den Einfluss von farbigen Nebenreizen auf den Wettstreit. *Pflüg. Arch. ges. Physiol.* **208**, 393–407.

GELLHORN, E. & SCHÖPPE, C. (1925*b*). Quantitative Untersuchungen über den Wettstreit der Sehfelder. V. Über den Einfluss der Beleuchtungsstärke sowie der Adaptation auf die Grösse der Wettstreitquotienten. *Pflug. Arch. ges. Physiol.* **208**, 408–17.

HOCHBERG, J. (1964). A theory of the binocular cyclopean field: On the possibility of simulated stereopsis. *Percept. mot. Skills* **19**, 685.

HUBEL, D. H. & WIESEL, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* **148**, 574–91.

HUBEL, D. H. & WIESEL, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–54

KAKIZAKI, S. (1960). Binocular rivalry and stimulus intensity. *Jap. psychol. Res.* **2**, 94–105.

KAPLAN, I. T. & METLAY, W. (1964). Light intensity and binocular rivalry. *J. exp. Psychol.* **67**, 22–6.

KAUFMAN, L. (1963). On the spread of suppression and binocular rivalry. *Vision Res.* **3**, 401–15.

LANSING, R. W. (1964). Electroencephalographic correlates of binocular rivalry in man. *Science* **146**, 1325–7.

LEVELT, W. J. M. (1965*a*). Binocular brightness averaging and contour information. *Brit. J. Psychol.* **56**, 1–13.

LEVELT, W. J. M. (1965*b*). *On Binocular Rivalry*. Soesterberg: Rep. Institute for Perception RVO-TNO; unpublished doctoral thesis, University of Utrecht.

MULL, H. K., ARMSTRONG, &. G TELFER, B. (1956). The effect of brightness upon reversible perspectives and retinal rivalry. *Amer. J. Psychol.* **69**, 123–5.

OGLE, K. N. (1950). *Researches in Binocular Vision.* Philadelphia and London: Saunders.

OGLE, K. N. (1961). Foveal contrast thresholds with blurring of the retinal image and increasing size of test stimulus. *J. opt. Soc. Amer.* **51**, 862–9.

ROELOFS, C. O. & ZEEMAN, W. P. C. (1919). Über den Wettstreit der Konturen. *v. Graefes Arch. Ophthal.* **99**, 79–104.

VERHOEFF, F. H. (1935). New theory of binocular vision. A.M.A. *Arch. Ophthal.*, **12**, 151–75.

WILDE, K. (1938). Figur und Fläche im Wettstreit. *Psychol. Forsch.* **22**, 26–58.

(*Manuscript received* 19 *March* 1965)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 239–253
Printed in Great Britain

EEG CORRELATES OF THRESHOLD REGULATION AS A FUNCTION OF STIMULUS WAVELENGTH: A COMPARISON BETWEEN NORMAL SUBJECTS AND PSYCHIATRIC PATIENTS

BY N. F. DIXON
Department of Psychology, University College, London, W.C. 1

Results of a previous study (Dixon & Lear, 1963, 1964) suggested the perceptual defence effect may be attributable to changes in pre-awareness levels of cortical activation. The present study looks into this apparent relation between threshold and activation by making use of the fact that the direction of threshold change, under conditions of sympathetic activity, appears to depend upon the wavelength of the stimulus test light. The results, using monochromatic light, suggest that the relation between threshold and activation, in normals and patients, may be less simple than previously supposed.

In an attempt to resolve the perceptual defence controversy by using a method which circumvented the criticisms (Dixon, 1958; Goldiamond, 1958; Eriksen 1958) levelled at earlier studies in this area it has been demonstrated (Dixon, 1960; Dixon & Haider, 1961) that the visual threshold for a spot of light presented to one eye may be significantly altered by the subthreshold presentation of emotive words to the other eye. A further investigation (Dixon, 1960) indicated that the site of such threshold changes was in all probability central rather than peripheral.

On the strength of these data a neurophysiological model was proposed to explain the observed phenomena. Based on ideas suggested by Samuels (1959) and the neurophysiological studies of Moruzzi & Magoun (1949), Jasper (1949), Ingvar & Hunter (1955), it incorporated two important findings: (1) that awareness of a visual stimulus depends upon the coincidence, in cortical reception areas, of impulses arriving over the specific sensory pathways and those arriving from upward discharge of the reticular activating system; (2) that the cortex by acting downwards on the R.A.S. can regulate its own level of activation.

Embodying these principles the model proposed that so-called perceptual defence and kindred phenomena depended upon the facilitation or inhibition of the reticular activating response as a function of a pre-awareness discrimination of inputs at a cortical level, from which it would follow that a relation should be found between thresholds for neutral and emotive words and the pre-awareness electroencephalogram (EEG).

A test of this hypothesis* (Dixon & Lear, 1963) did in fact provide data favourable to the model. Subjects with raised thresholds for emotive words showed sustained alpha abundance prior to awareness of the stimulus, while those with lowered thresholds showed a corresponding fall in alpha abundance. An analysis of slow wave components of the EEG, recorded under similar conditions (Dixon & Lear, 1964), showed that over all subjects the pre-awareness incidence of theta rhythm bore a U-shaped relation to threshold, a finding consistent with the curvilinear relation between anxiety and threshold reported by other workers (Brown, 1961).

---

* Since this paper was written Barratt & Beh (1964) and Barratt & Herd (1964) have shown that the alpha blocking response may be differentially conditioned to one of a pair of subliminal stimuli. This evidence for discrimination between subliminal stimuli with consequent control, by classical conditioning, over alpha activity provides considerable support for the findings by Dixon & Lear.

While suggestive of a simple and comprehensible relation between anxiety, activation and threshold these data did however leave open the question of causal direction. It could be argued on the one hand that, since threshold level was defined in terms of the time taken to perceive a slowly increasing visual stimulus, high thresholds would allow longer time for synchronous activity to build up. On the other hand, it could be the case that the alpha and theta phenomena merely reflected changes in autonomic activity occasioned by the stimulus and were not necessarily features of the threshold regulatory response.

To try to answer these questions, use was made in the present investigation of three phenomena concerned with the relation between thresholds, anxiety and the wavelength of light. The first is that sympathetico mimetic drugs (Kravkov, 1941) and emotionally disturbing stimuli (Dixon, 1960) decrease visual sensitivity for red light and increase it for green. The second is that the juxtaposition of a red stimulus and a meaningful mildly emotive stimulus (Drechsler, 1960*a*, *b*) evokes signs of emotional disturbance. The third is that light at the red end of the spectrum elicits physiological changes attributable to increased sympathetic activity (Birren, 1959).

The rationale for using red and green light, in the work that is reported here, was based on these data. If variations in the visual threshold depend solely upon preceding changes in levels of cortical activation (alpha abundance) then variations in threshold associated with changes in the wavelength of the exciting stimulus should be accompanied by corresponding changes in alpha abundance; in other words red and green should behave like white light. If, on the other hand, the EEG correlates of threshold regulation observed in the previous research really reflected concomitant changes in autonomic arousal, not necessarily related to threshold variation, then high thresholds for items presented in red light and low thresholds for those in green light should both follow upon reduced alpha abundance. Conversely, the reverse of these relations should hold for low thresholds for red and high thresholds for green light.

In addition to a re-examination of earlier findings by using light of different wavelengths the present investigation also involved a comparison between normal subjects and certain classes of psyhiatric patients. In a previous investigation (Dixon & Lear, 1962) it had been found that paranoid schizophrenics and people presenting depressive symptoms appeared to differ from normals in their manner of controlling the reception of affectively loaded material.

METHOD

General hypotheses

(1) Irrespective of stimulus wavelength an inverse relation will hold between threshold and the pre-awareness activation level.

(2) Significant differences in the EEG/threshold relation will occur as between normals and patients.

Subjects

The thirty-eight subjects were distributed as follows:

		Neurotic		Psychotic	
	Normal	Depression	Other	Depression	Schizophrenic
Male	3	3	2	3	7
Female	6	3	4	3	4

The normals were either hospital nursing staff or university students. The abnormal group were voluntary patients in hospital. (Diagnostic categories were assigned after interview by two and sometimes by three psychiatrists.)

Apparatus and stimulus presentation and recording methods

As in the previous studies, EEG were continuously recorded, from the dominant side, while the supine subject watched a dimly illuminated screen, upon which words were displayed at intensities which increased at 0·02 ft, lamberts/sec from zero to threshold brightness.

Threshold levels, the intensity at which the subject could first discern something on the screen, were indicated by a bell push in the subject's right hand.

EEG recordings were made, using 'Grass' needle electrodes, by an 'Offner Type T transistorized Dynograph'. In addition to the ink write-out they were fed simultaneously to a multi-track Fenlow magnetic tape recorder. This machine accepts EEG data by pulse width modulation of the input. On playback the EEG is demodulated to its original form.

The stimulus material presented during the EEG recording consisted of one emotional and one neutral word equated for length and frequency. For half the subjects the words were selected on the basis of results from a word association test. For the other half the two words 'CANCER' and 'DIVERS' were used as the emotive and neutral items respectively; CANCER was established as emotive on the basis of results from previous experiments (Dixon, 1960; Dixon & Haider, 1961; Coles, 1964). Each word was presented four times to each subject, twice in red light and twice in green light, subjects being assigned randomly to a counter-balanced design. Stimulus presentations and subjects' responses were recorded both on the paper write-out, and on one of the magnetic tape tracks. For both paper and tape an arbitrary code was chosen—so many deflexions in a particular channel on the paper and so many pulsed tones of different frequencies on the audio tape track.

Controls

To reduce the risk of artifacts arising from experimenter bias or the procedures used three kinds of controls were used. (1) The selection of subjects, recording of EEG's, stimuli presentation, and subsequent data analysis (replaying of the magnetic tapes) were carried out, respectively, by four different assistants. No one of these knew all the information needed to bias the results in any particular direction. (2) Possible contamination of the EEG data by muscle or other artifact was reduced by h.f. filters in the recording system and a C.R.O. monitor in the playback system. (Out of forty subjects tested two were rejected prior to data analysis because of EEG contamination by a.c. hum, due presumably to unduly high scalp resistance.) (3) Discrepancies between successive readings of the same tape by different assistants were checked by sampling the critical intervals over much longer epochs. Over a longer epoch the sampling error, due to loss of EEG information during the reading time, is eliminated.

Procedure

Following administration of the Ishihara test for colour blindness, electrodes were placed for bipolar recording from posterior to mid, parietal and temporal, positions. In an otherwise darkened room the projection screen above the subject's head was adjusted to a constant brightness of 0·2 ft lamberts. After a period for adaptation during which a trial EEG was recorded, the bell push was placed in the subject's hand. The subject was then given the following

Table 1 (*a*). *Mean alpha abundance and mean threshold for a 'red' stimulus (first presentation)*

		Pre-stimulus alpha. Means of 4 epochs		Stimulus alpha				Threshold (sec)	
				E_1+E_2		N_1+N_2			
	Subject	E_1+E_2	N_1+N_2	1st	2nd	1st	2nd	E_1+E_2	N_1+N_2
Threshold (E < N) and α(E > N)	NA (d) 23	81·0	82·7	78·5	81·0	77·5	77·5	31·5	32·7
	NA (d) 21	64·0	71·5	111·3	103·9	103·0	98·8	43·5	50·0
	PA (d) 29	79·2	75·8	62·0	68·1	72·6	56·5	40·0	51·0
	Mean	74·7	76·7	83·9	84·3	84·3	77·6	38·1	44·5
Threshold (E $\geqslant$ N) and α(E < N)	NA 22	201·5	181·5	198·0	177·0	184·3	194·1	65·5	65·5
	NA (d) 24	141·1	96·8	74·9	70·4	79·1	95·1	77·5	39·0
	Mean	157·8	139·1	136·4	123·7	131·7	144·6	71·5	52·2
Threshold (E < N) and α(E < N)	N 8	76·6	78·6	76·5	65·5	89·6	82·5	30·5	67·0
	N 7	139·8	150·7	143·2	95·0	153·0	104·0	32·0	39·0
	NA (d) 25	153·8	92·2	102·2	130·2	130·5	132·5	48·0	51·2
	NA (d) 28	171·2	137·0	108·0	101·5	108·0	211·5	23·0	28·0
	PA (d) 27	106·1	116·6	109·6	126·0	124·3	114·6	24·7	28·5
	Mean	129·5	115·0	107·9	103·6	121·1	129·0	31·6	42·7
Threshold (E > N) and α(E > N)	N 5	201·0	172·0	173·0	195·0	105·0	174·0	78·5	35·2
	N 6	111·6	113·0	115·5	108·9	88·3	82·6	52·3	30·4
	N 9	184·0	219·4	236·1	194·3	173·0	248·5	40·5	35·0
	NA 17	139·7	133·7	153·0	156·3	150·7	149·0	46·5	45·0
	NA 18	91·6	85·5	103·3	97·8	95·3	89·6	25·0	18·5
	SPA 19	63·6	47·8	66·7	50·3	55·2	54·0	25·5	23·0
	SPA 20	68·7	64·2	66·6	62·7	67·0	59·0	48·5	24·5
	SPA 26	109·7	101·7	102·6	96·5	94·5	98·2	45·0	40·5
	Mean	121·2	117·2	127·1	120·2	111·1	119·4	45·2	31·5

Abbreviations. Stimuli: emotive (E_1, E_2), neutral (N_1, N_2). Subjects: normal (N), neurotic (NA), psychotic (PA), schizophrenic (SPA), depressive (d).

Table 1(*b*). *Mean alpha abundance and mean threshold for a 'green' stimulus (first presentation)*

		Pre-stimulus alpha. Means of 4 epochs		Stimulus alpha				Threshold (sec)	
				E_1+E_2		N_1+N_2			
	Subject	E_1+E_2	N_1+N_2	1st	2nd	1st	2nd	E_1+E_2	N_1+N_2
Threshold (E > N) and α(E < N)	N 1	83·4	99·0	79·5	87·2	92·5	89·5	24·5	21·2
	NA 7	48·5	50·5	46·5	51·0	55·5	73·5	15·5	14·0
	PA(d) 8	140·3	134·1	132·5	134·1	142·5	131·0	44·0	32·5
	SPA 10	89·4	88·8	90·0	96·2	94·7	97·5	82·5	36·7
	Mean	90·4	93·1	87·1	92·1	96·3	97·8	41·6	26·1
Threshold (E < N) and α(E > N)	N 3	91·1	64·8	97·2	87·7	73·7	76·5	23·0	23·1
	SPA 14	44·7	48·6	51·0	41·2	47·8	43·4	21·5	40·0
	SPA 2	118·8	106·2	146·7	119·7	84·7	65·5	37·7	38·6
	Mean	84·9	73·2	94·9	82·9	68·7	61·8	27·4	33·9
Threshold (E < N) and α(E < N)	N 4	164·0	215·6	163·0	185·0	188·2	206·5	18·5	21·0
	NA 16	127·5	128·5	116·1	117·6	121·1	125·1	44·0	47·2
	PA 11	114·5	93·2	100·0	108·0	107·0	103·0	29·5	35·0
	PA 12	70·0	74·2	72·8	66·7	69·7	76·7	70·5	87·0
	PA(d) 13	123·6	117·7	68·5	145·0	171·5	150·0	22·0	28·0
	SPA(d) 15	87·7	98·1	88·8	69·8	84·5	86·8	24·5	48·5
	Mean	114·6	121·2	101·5	115·3	123·6	124·7	34·8	44·4
Threshold (E > N) and α(E > N)	N 2	114·7	106·0	112·0	129·0	105·0	124·0	18·0	15·5
	NA(d) 1	171·7	173·5	194·2	144·2	167·5	169·7	54·2	37·0
	NA 9	95·5	70·0	76·8	69·1	78·8	64·5	42·0	37·5
	SPA 6	241·7	256·5	277·5	240·9	247·0	249·9	69·0	27·0
	PA(d) 3	168·2	149·4	179·3	192·0	163·8	198·8	35·7	34·0
	SPA 5	100·0	107·6	130·6	110·2	92·1	73·2	85·5	38·5
	SPA 4	69·7	64·1	69·1	66·0	63·6	64·6	30·0	27·5
	Mean	137·4	132·4	148·5	135·9	131·1	134·9	47·8	31·0

instructions. 'We are going to project some words at very low intensities. As soon as you think you can see something appearing on the screen press the button once. In between times just relax but keep watching the screen.'

A practice stimulus word was presented and further instructions given if necessary. The various experimental stimuli were then presented according to the assigned order. The periods of time between stimulus presentations were randomly varied in length from fifteen to thirty seconds. After half the presentations, the colour filters were changed from red to green (or vice versa) without the subject's knowledge.

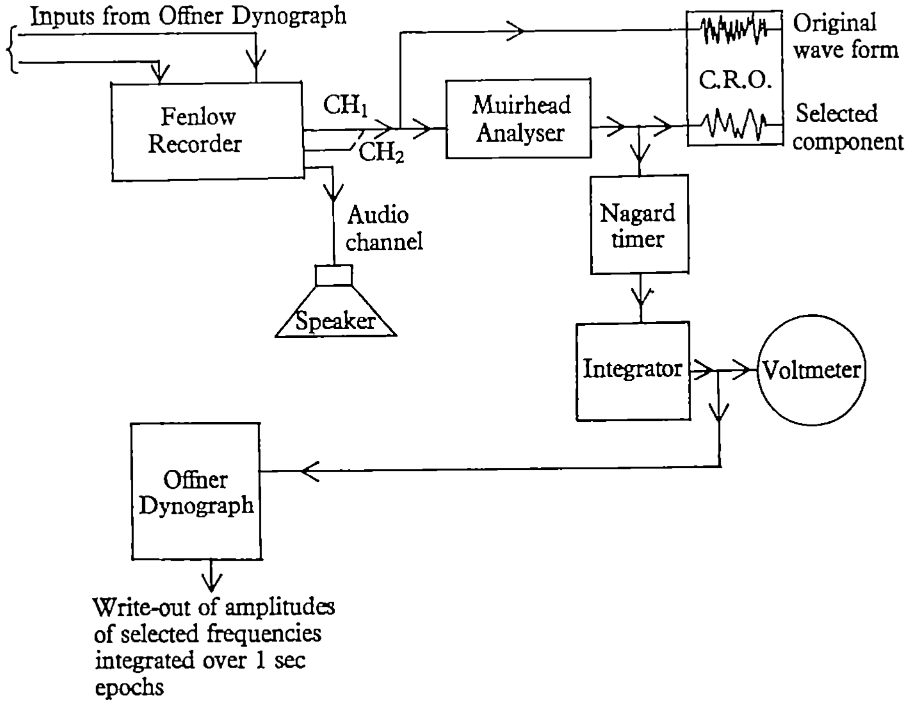


Fig. 1. Lay-out of the system used for recording and analysing the data.

Results

The arrangement shown in Fig. 1 was used for analysis of the data. By reading off the integrator, which sums amplitudes over any interval determined by the timer, it is possible to quantify the abundance of any component of the EEG that has been selected by the analyser, over any desired length of epoch. Since the EEG is 'locked' to the audio record of stimulus or response events it is then possible to examine the EEG data within each of the experimental and control conditions.

Table 1 contains the scores for alpha abundance and threshold in respect of all subjects. The first two columns show the mean alpha (integration readings) recorded during the four epochs prior to presentations of emotive and neutral stimuli respectively. The next four columns show mean alpha recorded in the first and second halves of the emotive and neutral stimulus trials respectively (i.e. between stimulus onset and awareness of the stimulus). The last two columns show the mean thresholds in seconds for the two sorts of stimulus material. Subjects have been placed in this table under the four possible relations between alpha and threshold for the two sorts of stimulus material. It is apparent that the distribution of subjects over these four possibilities provides support for our earlier finding that relatively higher thresholds are concomitant with relatively more alpha in the pre-stimulus-awareness EEG. The actual numbers of subjects falling into each category are shown in Table 2. The relation is significant for the 'red' group ($P < 0{\cdot}05$) and for the two groups combined ($P < 0{\cdot}02$; sign test one tail probabilities in both cases). On the face of it they suggest that the relation between EEG and visual threshold is independent of stimulus wavelength.

Table 2. *Classification of subjects by α-abundance and threshold for emotive (E) and neutral (N) stimuli*

'Green' presentation			'Red' presentation		
	Alpha abundance			Alpha abundance	
Threshold	$\alpha_E > \alpha_N$	$\alpha_E < \alpha_N$	Threshold	$\alpha_E > \alpha_N$	$\alpha_E < \alpha_N$
E > N	7	4	E > N	8	2
E < N	3	6	E < N	3	5

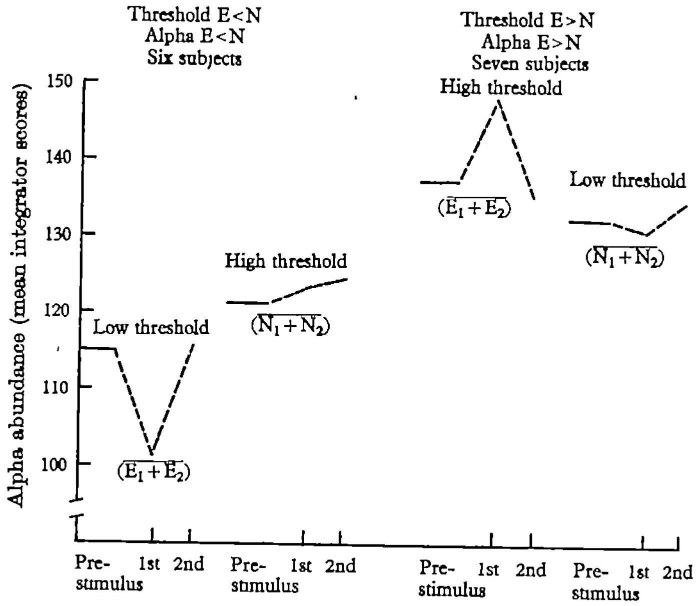


Fig. 2. Alpha abundance as a function of threshold for emotive (E) and neutral (N) words shown in 'green' light (*first* colour presented); ——, means of four epochs before stimulus presented; ---- means of first and second halves of each stimulus trial.

This conclusion is supported by comparisons of alpha abundance before and *after* the stimulus was presented (see Figs. 2 and 3).

Alpha decreased for low threshold words and increased for high ones in both 'red' and 'green' stimulus conditions. These changes appeared to occur within the first half of each stimulus trial. That 'red' evidently had a greater effect than 'green' is supported by the fact that whereas in the group of subjects presented with a 'red' stimulus the correlation between awareness threshold and alpha abundance during the first half of each stimulus trial was significantly different from zero ($\rho = 0{\cdot}46$, Table 3 (*a*). *Mean alpha abundance mean and threshold for a 'green' stimulus (second presentation)*

		Pre-stimulus alpha. Means of 4 epochs		Stimulus alpha				Threshold (sec)	
				E_1+E_2		N_1+N_2			
	Subject	E_1+E_2	N_1+N_2	1st	2nd	1st	2nd	E_1+E_2	N_1+N_2
Threshold (E > N) and α(E < N)	N 7	201·1	149·4	124·3	130·1	136·5	131·0	35·0	29·0
	N 9	143·1	122·4	135·2	132·7	171·5	204·0	18·5	13·5
	NA 17	151·2	145·2	153·2	159·2	152·4	171·1	41·0	36·5
	Mean	165·0	139·0	137·5	140·7	163·5	168·7	31·5	26·3
Threshold (E < N) and α(E > N)	NA(d) 25	75·5	79·7	72·5	97·0	66·7	91·2	19·8	23·2
	NA 22	198·0	205·2	207·4	192·5	180·8	177·2	50·5	51·0
	NA(d) 28	202·1	159·5	193·0	193·0	145·0	156·5	8·5	18·0
	PA(d) 27	109·4	113·7	104·2	101·5	103·5	92·5	21·5	24·0
	NA 18	83·6	89·2	96·1	84·2	83·2	80·5	21·5	22·5
	NA(d) 23	92·6	90·6	89·5	82·8	86·0	83·0	24·0	26·5
	N 8	92·6	72·0	73·0	86·0	67·2	66·5	15·2	17·5
	Mean	120·2	116·6	119·4	119·6	104·6	106·8	23·0	26·1
Threshold (E < N) and α(E < N)	PA(d) 29	53·4	74·0	84·5	109·5	111·7	95·8	41·7	52·5
	NA(d) 21	75·0	61·7	66·2	94·3	75·8	91·9	27·0	31·5
	NA(d) 24	68·0	64·7	49·7	31·0	53·7	46·2	24·5	28·5
	N 6	113·0	110·5	109·1	110·1	129·2	99·6	30·8	32·6
	Mean	77·3	77·7	77·4	86·2	92·6	83·4	31·0	36·3
Threshold (E > N) and α(E > N)	N 5	181·6	158·9	183·5	131·1	130·7	143·7	30·0	23·0
	SPA 26	114·7	113·2	136·7	101·3	105·7	77·2	64·0	34·5
	SPA 20	65·5	71·4	65·1	75·1	59·6	69·5	95·0	77·5
	SPA 19	62·4	55·1	53·7	66·1	49·6	58·7	18·0	16·0
	Mean	106·0	99·6	109·7	93·4	86·4	87·2	51·7	37·7

Table 3(*b*). *Mean alpha abundance and mean threshold for a 'red' stimulus (second presentation)*

		Pre-stimulus alpha. Means of 4 epochs		Stimulus alpha				Threshold (sec)	
				E_1+E_2		N_1+N_2			
	Subject	E_1+E_2	N_1+N_2	1st	2nd	1st	2nd	E_1+E_2	N_1+N_2
Threshold (E > N) and α(E < N)	N 4	269·4	223·8	194·6	219·8	240·5	218·5	29	27 5
	NA 7	44·1	76·1	42·7	44·8	73·7	79·0	32	19·7
	NA 9	104·5	79 5	86·5	63·0	97·0	97·0	35	33·0
	SPA 2	78·5	92·5	69·6	82·2	88·6	65·3	125	105·0
	PA(d) 8	112·0	140·4	110·4	100·3	151·2	165·4	91·5	89·5
	SPA 11	120·2	108·7	113·7	81·3	87·3	122·3	59·0	27·0
	SPA 12	71·0	69·7	74·7	77·7	85·8	74·3	96 5	74·0
	PA(d) 13	130·8	103·8	99·3	149·0	153·7	139·1	47·0	37·5
	SPA(d) 15	108·0	104·7	107·9	92·8	118·0	86·2	46 0	42·0
	PA(d) 3	201·6	227·0	160·5	131·6	146·3	199·5	47·0	42·0
	Mean	124·0	122·6	106·0	104·2	124·2	124·7	60·8	49·7
Threshold (E < N) and α(E > N)	N 1	88·2	153·2	93·5	88·2	83·7	93·5	26·0	28·0
	N 2	99·7	99·8	125·0	119·4	103·5	105 5	23·5	30·5
	N 3	82·6	45·4	76·5	68 2	58·2	51·0	32·0	34·5
	NA 16	132·2	117·7	154·0	160·5	121·0	138·0	128·0	146·0
	SPA 4	64·2	74·0	78 3	72·3	75·1	73 6	29·5	30·5
	SPA 5	212·2	251·8	278·3	224·6	262·3	232·3	47·5	66·0
	SPA 6	276·2	291·5	283·3	282·1	297·9	251·2	48·0	54·5
	Mean	136·5	147·6	155·5	145·0	143 1	135·0	47·8	55·7
Threshold (E < N) and α(E < N)	NA(d) 1	177·0	161·4	142·3	144·0	171·1	184·9	42·5	64·0
	SPA 14	48·7	42·0	44·3	39·7	43·7	41·6	28·0	37.0
	Mean	112·8	101·7	93·3	91·8	107·5	113·2	35·2	50·5
Threshold (E > N) and α(E > N)	SPA 10	94·1	97·2	126·0	114·0	106·0	108·4	64	44

Abbreviations. Stimuli: emotive (E), neutral (N). Subjects: normal (N), neurotic (NA), psychotic (PA), schizophrenic (SPA), depressive (d).

$P < 0{\cdot}05$), for those having the 'green' stimulus the correlation fell short of significance ($\rho = 0{\cdot}27$). Contrary to expectation, no significant difference was found between normals and patients.

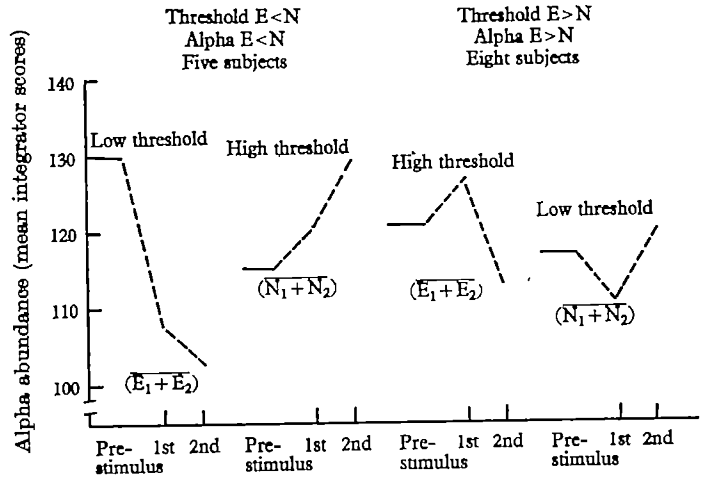


Fig. 3. Alpha abundance as a function of threshold for emotive (E) and neutral (N) stimulus words shown in 'red' light (*first* colour presented); ——, means of four epochs before stimulus presented; ----, means of first and second halves of each stimulus trial.

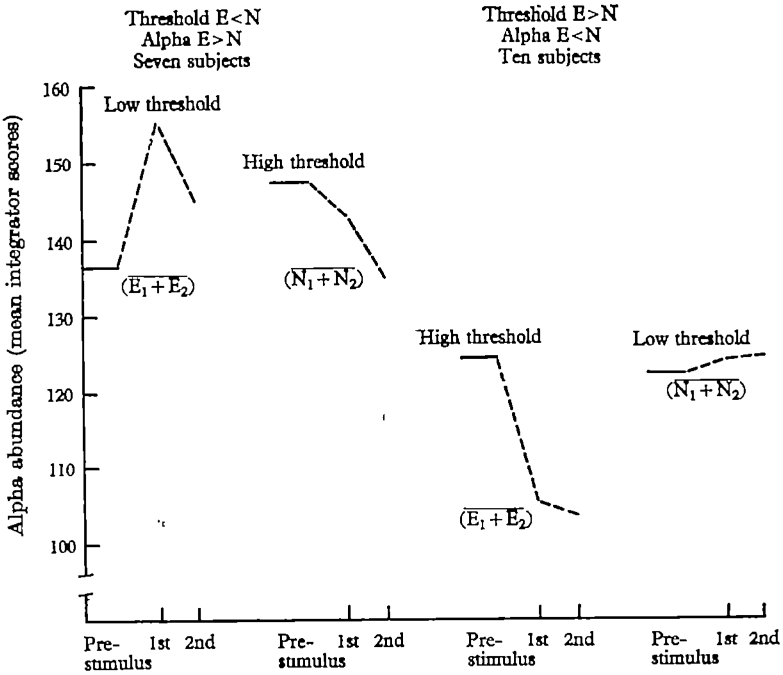


Fig. 4. Alpha abundance as a function of threshold for emotive (E) and neutral (N) stimulus words shown in 'red' light (*second* colour presented); ——, means of four epochs before stimulus presented; ----, means of first and second halves of each stimulus trial.

Second colour presentation

After two presentations of each of the stimulus words in one colour, all subjects were given identical treatment with the other colour. By this time it may be assumed that all subjects would have become aware of the nature of the four stimulus words. The results are continued in Table 3 and Fig. 4.

Table 4. *Relation between α-abundance and threshold in the 'red' group*

	Alpha abundance	
Threshold	$\alpha_E > \alpha_N$	$\alpha_E < \alpha_N$
E > N	1	10
E < N	7	2

Table 5. *Relation between threshold and type of subject for the seventeen subjects in the leading diagonal cells of Table* 4

	Thresholds	
	E > N	E < N
Patients	9	4
Normals	1	3

Table 6. *Relation between type of visual stimulus and tendency to inverse relation of threshold and α-abundance*

	Threshold E > N and alpha E < N	Threshold E < N and alpha E > N
Red	9	4
Green	1	6

It appears that while the relation between threshold and alpha abundance under stimulation with 'green' light was without significance, that under 'red' showed a striking inversion of the relation found during the first presentation. With three exceptions all subjects showed relatively more alpha prior to awareness of words for which they had low thresholds and less alpha prior to awareness of words for which they had high thresholds. The distributions are shown in Table 4. The Fisher–Yates test shows the relation to be significant ($P < 0{\cdot}01$, two tail test).

Of the seventeen subjects who contributed to this inverse relation between alpha abundance and threshold, patients tended to have raised, and normals to have lowered thresholds for 'emotive' stimuli (see Table 5). In this respect patients in the 'red' groups tended to differ from those in the equivalent 'green' group (see Table 6). These results fall just short of that required for significance (at $P < 0{\cdot}05$, two tail Fisher–Yates test).

A second finding was that, within the patient group, depressives tended to have raised thresholds for emotive words when presented in red light, and lowered thresholds when presented in green (see Table 7; $P < 0{\cdot}05$ for two tail Fisher–Yates test).

Thresholds

A significant relation between pre-awareness alpha abundance and threshold for a word does not necessarily imply any pre-awareness discrimination of the latter *unless* some relation can be found between the nature of the word and the subject's threshold. Without this relation, the most parsimonious explanation of the results would be merely that, when activation level is low (for any reason whatever), thresholds will be high and vice versa. In the present instance there are two kinds of evidence which support the view that some discrimination of the meaning of the stimulus did, in fact, occur prior to awareness.

In the first place there appears to be a curvilinear relation between threshold and emotionality of the stimulus (Table 8). This tendency for subjects to have higher or lower thresholds for E than for N words is consistent with data from previous research.

Table 7. *Relation between threshold and type of stimulus for depressive patients*

	Thresholds	
Stimulus	E > N	E < N
Red	4	1
Green	0	7

Table 8. *Relation between threshold and emotionality of stimulus*

		Mean thresholds for subjects with			
		E thresholds > N thresholds		E thresholds < N thresholds	
1st presentation	Green	44·7	28·5	31·1	39·1
	Red	58·3	41·8	34·8	43·6
2nd presentation	Green	41·6	32·0	27·0	31·2
	Red	62·4	46·8	41·5	53·1
	Mean	51·7	37·2	33·6	41·7

The second piece of evidence for discrimination is that on the second presentation twenty-six of the thirty-eight subjects showed higher thresholds for the first E word than for the first N word. By the sign test this finding has a two-tail probability of 0·033.

On the basis of these data we would tentatively assume that the relation between alpha abundance and threshold depends upon some prior discrimination of stimulus meaning.

DISCUSSION

This investigation was designed to test a hypothesis arising from a neurophysiological model for threshold regulation. According to the model, raised or lowered thresholds for emotive material result from pre-awareness changes in cortical activation. While the data from the first part of the experiment support this hypothesis, and are in line with the results from a previous study using white light, those from the second part do not. So far as red light is concerned the model is either wrong or incomplete. In view of the considerable evidence for a relation between cortical activation and excitability (Jasper, 1949; Murphy & Gellhorn, 1945; Dixon & Lear, 1963) to what then can these results be attributed?

The following considerations seem to be relevant to a provisional explanation. Assuming from the present data and the research mentioned in the introduction that there *is* a relation between the wavelength of light and arousal, this might well be encompassed by the sort of photo-neuroendocrine effects recently described by Hollwich (1964) and others. These effects, which may depend upon the existence of extra-visual retino-hypothalamic pathways as distinguished in various subhuman species (Jacobs & Morgane, 1964; Feldman, 1964), include the unique influence of red light upon hypothalamic responsiveness (Benoit, 1964). In this connexion, red light has been shown to evoke sympathetic excitation* (Birren, 1959), and emotional disturbance (Drechsler, 1960, 1961), in human subjects; moreover, in the first part of the present experiment a 'red' stimulus appeared to evoke a stronger relation between alpha and threshold than did a 'green' stimulus. (Walter & Walter (1949), have also reported the 'dramatic' effect upon the EEG of photic driving with a *red* light; 'the effect includes an increased incidence of spikes and high frequency components'.)

This sympathetic activity might be expected to evoke cortical activation with a reduced incidence of alpha, and E. Gellhorn (personal communication) has suggested that the raising or lowering of human visual thresholds may reflect altered states of sympathetic/parasympathetic balance. Certainly it would seem that hypothalamic excitation in cats enhances cortical sensitivity for specific (optic and acoustic) stimuli largely through an increase in the threshold (caudate nucleus) of the unspecific inhibitory recruiting system (Gellhorn & Loofbourrow, 1963). But from the work of Kravkov & Galochkina (1947) it would also seem that sympathetico-mimetic agents and therefore, presumably, sympathetic activity decrease peripheral sensitivity for red light. That central processes may play a part in determining retinal sensitivity is also suggested by the work of Granit (1955*a*, *b*) on antidromic factors in vision. In other words, the peripheral and central aspects of autonomic activity could together account for the apparently anomalous finding that raised thresholds for emotive words, presented in red light, followed upon reduced alpha abundance.

Caution is needed in attaching any weight to these conjectures. Other explanations are possible. The results are not as clear-cut as one might wish, and again, the data from a previous study (Dixon, 1960) support the view that differential changes in colour thresholds are much more likely to be *central* in origin, a view favoured by H. Davson (personal communication). Clearly, too, it will be necessary to examine the question of possible interactions between the different colour presentations that have been used.

---

* Since this paper was written a number of experiments at University College by Dixon and Maxeiner investigating the hypothesis that red light 'acts' as a thermal stimulus suggest that, during the course of thermoregulation to changes in ambient temperature, optic irradiation by red light (infra-red free) always promotes peripheral vasoconstriction, i.e. red acts sympathetically and does not appear to be 'misinterpreted' as a thermal stimulus.

This research was made possible by a grant from the Mental Health Research Fund. I should like to express my gratitude to Dr T. E. Lear, Mr J. Chambers and the staff of the EEG Department of University College Hospital for their generous assistance.

REFERENCES

BARRATT, P. E. H. & HERD, J. M. (1964). Subliminal conditioning of the alpha rhythm. *Austr. J. Psychol.* **16**, 9–19.

BARRATT, P. E. H. & BEH, H. C. (1964). Subliminal discrimination and the concept of vigilance. *Austr. J. Psychol.* **16**, 107–19.

BENOIT, J. (1964). The role of the eye and of the hypothalamus in the photo-stimulation of gonads in the duck. *Ann. N.Y. Acad. Sci.* **117**, 204–15.

BIRREN, F. (1959). The effect of colour on the human organism. *Amer. J. occup. Therap.* **13**, 125–9.

BROWN, W. P. (1961). Conceptions of perceptual defence. *Brit. J. Psychol.* (Monogr. Suppl.), no. 35.

COLES, E. M. (1964). The effects of anxiety on the brightness threshold. Unpublished Ph.D. thesis, University of London.

DIXON, N. F. (1958). Apparent changes in the visual threshold as a function of subliminal stimulation. *Quart. J. exp. Psychol.* **10**, 211–19.

DIXON, N. F. (1960). Apparent changes in the visual threshold: central or peripheral? *Brit. J. Psychol.* **51**, 297–309.

DIXON, N. F. & HAIDER, M. (1961). Changes in the visual threshold as a function of subception. *Quart. J. exp. Psychol.* **13**, 229–35.

DIXON, N. F. & LEAR, T. E. (1962). Perceptual regulation and mental disorder. *J. ment. Sci.* **108**, 356–61.

DIXON, N. F. & LEAR, T. E. (1963). Electroencephalograph correlates of threshold regulation. *Nature, Lond.*, **198**, 870–2.

DIXON, N. F. & LEAR, T. E. (1964). Incidence of theta rhythm prior to awareness of a visual stimulus. *Nature, Lond.*, **203**, 167–70.

DRECHSLER, R. J. (1960). Affect-stimulating effects of colours. *J. abnorm. soc. Psychol.* **61**, 323–8.

DRECHSLER, R. J. (1961). Subliminal influences upon affective responses to colour. *Amer. Psychologist*, **16**, 428.

ERIKSEN, C. W. (1958). Unconscious processes. In *Nebraska Symposium on Motivation*, ed. M. R. Jones. Lincoln: University of Nebraska Press.

FELDMAN, S. (1964). Visual projections to the hypothalamus and preoptic area. *Ann. N.Y. Acad. Sci.* **117**, 53–68.

GELLHORN, E. & LOOFBOURROW, G. N. (1963). *Emotions and Emotional Disorders.* New York: Harper and Row.

GOLDIAMOND, I. (1958). Indicators of perception. *Psychol. Bull.* **55**, 373–411.

GRANIT, R. (1955*a*). Brain control of the sense organs. *Acta Psychol., Amsterdam*, **11**, 117–18.

GRANIT, R. (1955*b*). Centrifugal and antidromic effects on ganglion cells of the retina. *J. Neurophysiol.* **18**, 388–411.

GRANIT, R. (1955*c*). *Receptors and Sensory Perception.* New Haven: Yale University Press.

HOLLWICH, F. (1964). The influence of light via the eyes on animals and man. *Ann. N.Y. Acad. Sci.* **117**, 105–31.

INGVAR, D. H. & HUNTER, J. (1955). Influence of visual cortex on light impulses in the brain stem of the unanaesthetized cat. *Acta physiol., Scand.* **33**, 194–218.

JACOBS, M. S. & MORGANE, P. J. (1964). Retino-hypothalamic connections in cetacea. *Nature, Lond.*, **203**, 778–80.

JASPER, H. (1949). Diffuse projection system of the thalamus. *E.E.G. clin. Neurophysiol.* **1**, 403–420.

KRAVKOV, S. V. (1941). Colour vision and the autonomic nervous system. *J. opt. Soc. Amer.* **31**, 335–7.

KRAVKOV, S. V. & GALOCHKINA, L. P. (1947). Effect of a constant current on vision. *J. opt. Soc. Amer.* **37**, 181–6.

MORUZZI, G. & MAGOUN, H. W. (1949). Brain stem reticular formation and activation of the E.E.G. *E.E.G. clin. Neurophysiol.* **1**, 455–73.

MURPHY, J. P. & GELLHORN, E. (1945). The influence of hypothalamic stimulation on cortically induced movements and on action potentials of the cortex. *J. Neurophysiol.* **8**, 339–64.

SAMUELS, I. (1959). Reticular mechanisms and behaviour. *Psychol. Bull.* **56**, 1–25.

WALTER, V. J. & WALTER, W. G. (1949). Central effects of sensory stimulation. *E.E.G. clin. Neurophysiol.* **1**, 57–86.

(*Manuscript received* 11 *January* 1965; *revised manuscript received* 26 *November* 1965)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 255–261
Printed in Great Britain

TWO TYPES OF SET AND THE GENERALIZATION OF PERCEPTUAL DEFENCE

By ANN TAYLOR
University of Leicester
AND DEREK W. FORREST
Trinity College, Dublin

The present paper takes as its starting point the view that the apparent generalization of perceptual defence is to be attributed to an interruption of the observer's set. Two possibilities with regard to the nature of this set are examined, and the experimental results clearly support one of them: the view that 'a set to see neutral words' is operative. Difficulties in understanding this formulation are discussed.

McGinnies & Sherman (1952) demonstrated that a 'perceptual defence effect' can be obtained in an experimental situation in which words are presented tachistoscopically in pairs. A *pretask* word is first presented for 2 sec, at a brightness well above that required for identification; it is then immediately replaced by a rapidly exposed *task* word which the subject must attempt to recognize. Recognition is slower for task words that follow pretask words of strong emotional connotation than for task words that follow neutral pretask words. This result was explained by McGinnies and Sherman in terms of the generalization, or spread in time, of aversive reactions set up by presentation of the emotional pretask words. Furthermore, their evidence was not susceptible to an interpretation in terms of response readiness as the task words themselves are not systematically different in this respect. The finding has been corroborated by other experimenters using subliminal presentation of the pretask words (Walters, Banks & Ryder, 1959).

Forrest, Gordon & Taylor (1965) have shown that the 'generalization of perceptual defence' can be produced when the emotional pretask words are replaced by nonsense words; they have also shown that when subjects are given practice pairs which include both neutral and emotional pretask material the effect does not occur. This evidence suggests that the effect may be due less to the operation of generalized aversive reactions than to the disruption of set produced by the appearance of incongruous pretask material. If this interpretation is correct a question arises as to the exact nature of the set involved.

Two possibilities suggest themselves. It may be, as Freeman (1954) and others have suggested, that subjects in the laboratory are normally set to perceive affectively neutral rather than emotional material. Alternatively, it is possible that subjects, either spontaneously or as a result of their experience with practice pairs, expect that the second word of a pair will be in some sense similar to the first, and thus that emotional task words will follow emotional pretask words; hence, when neutral words do in fact follow, they have higher recognition thresholds.

The present experiments were designed to distinguish between these two possibilities. Subjects were supplied with word lists from which they chose their answers.

Each list consisted of all the task words for a particular subject and also additional words, either affectively neutral or emotional. In one experiment subjects were given lists of neutral words only, while in another experiment they were given lists containing both neutral and emotional words.

If subjects have a set to perceive neutral rather than emotional words, this set would be reinforced by giving them a list of neutral words from which all correct identifications were to be drawn; but the set would be abolished by giving them a list containing emotional as well as neutral material. If, on the other hand, subjects have a set to perceive stimuli in pairs related in emotional tone, such a set would be abolished by giving them an entirely neutral list of possible responses but not by giving them a mixed list. Hence, if the generalization effect is found in the first experiment and not in the second, the disruption of a set to perceive neutral material would be indicated; if the effect occurs in the second experiment it should rather be ascribed to the operation of a set to perceive the second word of a pair as affectively related to the first.

The subjects in these experiments were men and women undergraduates at a university for both sexes. As the generalization of perceptual defence has so far been demonstrated within groups of American men college students (McGinnies & Sherman, 1952; Walters *et al.* 1959) and British women undergraduates from a college for women (Forrest *et al.* 1965), it was thought advisable to replicate the original procedure using subjects from the present population. Expt I of the present study is this replication.

METHOD

Subjects

Thirty men and thirty women undergraduates aged between 18 and 22 yr served as subjects, ten men and ten women being assigned at random to each of the three experiments. Some subjects were first-year students of psychology; the remainder had studied no psychology. All were unaware of the purpose of the investigation.

The stimulus words

The five-letter words of McGinnies and Sherman were used with some additional words, both neutral and emotional. Thirty-seven different words were used, twenty-seven of them being included as neutral and ten as emotional words. Twenty-two of the words were to be presented tachistoscopically; and the remaining fifteen were used in making up the word lists needed for Expts. II and III. The twenty-two words to be shown tachistoscopically were chosen afresh from the pool for each subject.

Neutral words: Abide, Baron, Brand, Cable, Creek, Dense, Dwarf, Flush, Frock, Glide, Hound, Jelly, Legal, Lemon, Lucky, Miner, Outer, Phone, Quest, Ranch, Rider, Scent, Stray, Thorn, Towel, Weave.

Emotional words: Balls, Belly, Bitch, Filth, Penis, Pubic, Raped, Semen, Urine, Whore.

All the neutral words were listed by Thorndike & Lorge (1944) with frequencies of usage between 18 and 28 per million. Of the emotional words, 'Balls', 'Bitch', 'Penis', 'Pubic', and 'Semen' do not appear in the count: 'Urine' is quoted at a frequency of 1 per million, 'Raped' and 'Whore' at 3 per million, 'Filth' at 4 per million, and 'Belly' at 10 per million.

Procedure

The procedure of McGinnies and Sherman was followed. Eleven pairs of words were presented, three in the practice trials and eight in the experiment proper. Of these eight pairs, four had an emotional pretask word and four a neutral pretask word. The pretask words of the practice trials and all the task words throughout the experiment were neutral. With these limitations, the words were randomly paired and ordered for each subject.

No list was given to subjects in Expt. I. Each word list used in Expt. II and Expt. III comprised the eleven words shown to the subject in the testing session and nine additional words. In Expt. II (the neutral list condition) the nine extra words were simply those remaining in the pool of neutral words. Subjects in Expt. III (the mixed-list condition) had lists in which the task words were supplemented by three additional neutral words chosen at random from the pool and the six emotional words which were not in use as pretask words.

Pretask and task words were separately projected by two 'Leitz Pradovit' projectors connected by a timing device. When the experimenter activated the timer the appropriate slides in both projectors slid into position, the word in the pretask projector being shown on the screen while the task word remained occluded by an electronic shutter. After 2 sec the timer automatically switched off, the slide in the pretask projector was removed, and the shutter of the task projector was released to permit a brief exposure of the second word. The brightness of each projector could be independently controlled by means of a variac transformer.

Subjects were tested individually by a woman experimenter in a sound-deadened room at one end of which a white screen was placed. The subject sat at a distance of 5 ft from the screen and to one side of the projectors and thus at a slight angle to the screen. The incident illumination was 1·5 ft-c. when the projectors were off. The projected words were 2 in. high and 9 in. wide.

Instructions

The following instructions were given to subjects in all three experiments: 'When I say "Ready" a word will appear on the screen and stay there long enough for you to recognize it easily. As soon as it disappears, another word will flash in its place, this time only for a very brief second, and I want you to try to tell me what this second word is.' Subjects in Expts II and III were then also told: 'The words you have to identify will all be taken from the words on this list. Perhaps you would like to read it through once to yourself.' Each subject was given his specially prepared list and allowed to read it through. In all three experiments the experimenter then continued: 'Is all that clear? Are there any questions? O.K., then we'll start with the first pair and see how it goes.'

The three practice pairs were then shown to the subject and his recognition threshold for the task words was roughly determined. When the subject failed to recognize a task word, both words of the pair were again shown at the same exposure (5 msec) but the brightness of the task projector was increased. For the main experiment the variac controlling the brightness was set initially at 40 V below the lowest threshold recorded in the practice trials. On the second and subsequent trials leading up to recognition the voltage was increased by 10 V steps; this large step-size seemed to be required by the rather high incident illumination. The brightness of the pretask projector was held constant at 180 V.

When all the task words had been correctly identified the subject was asked if there were any words on the list or if any words had been used in the experiment which he did not understand. He was also asked what he thought the experiment was about, what he had expected to see, and what he had thought when emotional words were presented.

RESULTS

Data from each experiment were analysed separately. The statistical procedure of McGinnies & Sherman (1952) was followed. The total numbers of trials were calculated for recognition of the four task words that followed neutral pretask words, and for recognition of the four task words that followed emotional pretask words. These totals were averaged for the twenty subjects in each experiment and t tests were used to compare the means in each case.

Table 1 gives these means for each experiment. It can be seen that in both Expts I and II a significant 'defence' effect was found in that subjects required more trials in order to recognize words correctly after emotional pretask words. In Expt. III subjects took rather fewer trials to recognize these words than words which followed neutral pretask words, but the difference between means does not approach significance. The t tests were supplemented by sign tests on the direction of differences between the totals for each subject. Table 2 gives the number of subjects in each experiment who required more exposures to recognize task words after emotional than after neutral pretask words and vice versa. The sign tests support the t tests and indicate a significant 'defence' effect in Expts. I and II and a non-significant difference in Expt. III. Hence it can be concluded that the generalization effect is obtained both when no response list is given and when the list available comprises only neutral words, but that the effect is abolished when emotional as well as neutral words appear in the list.

Table 1. *Differences in recognition thresholds for subjects in the three experiments: exposures needed to recognize task words* (*a*) *after neutral pretask words*, (*b*) *after emotional pretask words*

Expt.	n	(*a*) mean	(*b*) mean	S.E. diff.	D.F.	t	P
I	20	18·2	20·6	0·92	19	2·60	$< 0·02$
II	20	14·7	16·0	0·54	19	2·39	$< 0·05$
III	20	15·4	14·6	0·53	19	1·52	N.S.

Table 2. *Differences in recognition thresholds for subjects in the three experiments: number of exposures needed for recognition of task words* (*a*) *after neutral pretask words*, (*b*) *after emotional pretask words*

Expt.	n	No. of subjects for whom $a < b$	No. of subjects for whom $a > b$	Sign test P
I	20	14	4	$< 0·015$
II	20	15	3	$< 0·004$
III	20	6	10	$< 0·227$

Pre-recognition guesses

In all three experiments, sign tests reveal no significant differences between the numbers of hypotheses proposed by subjects on pre-recognition trials after neutral and after emotional pretask words.

It was possible to divide pre-recognition guesses into emotional and neutral categories and then to tabulate the frequencies with which these categories followed neutral or emotional pretask words. In Expt. I, seventy-one pre-recognition guesses were made, of which only four at the most could be considered emotional or sexual in connotation. The four were 'Love', 'Toilet', 'Crook', and 'Flies'. All these words were produced after emotional pretask words: 'Love' and 'Crook' after 'Whore', 'Toilet' after 'Belly', 'Flies' after 'Balls' (subsequently admitted by the subject, a woman, to denote trouser fastening). With so small a number it is of course impossible to draw any firm conclusion.

In Expt. II there were no emotional hypotheses, all those that were made being taken from the 'neutral' list given to each subject.

Expt. III (see Table 3) revealed a number of tendencies, but none unambiguously.

(*a*) More neutral than emotional words were used as pre-recognition hypotheses whatever the nature of the pretask word. Twelve subjects made more neutral than emotional pre-recognition guesses, while only three showed the reverse effect; but a result of this kind is to be expected on the basis of random guessing, since there were more neutral than emotional words in the list. The total numbers of neutral and emotional pre-recognition guesses were 25 and 11 respectively; this ratio corresponds closely to the 14:6 bias of the word lists themselves. (*b*) There is some indication that pre-recognition hypotheses tended to be of the same nature as the immediately preceding pretask word, i.e. more guesses fall into the neutral–neutral or emotional–emotional categories than into the neutral–emotional or emotional–neutral categories; but the tendency is not strong.

Table 3. *Number of pre-recognition guesses in Expt. III*

(N–N, neutral pretask word, neutral guess; E–E, emotional pretask word, emotional guess; N–E, neutral pretask word, emotional guess; E–N, emotional pretask word, neutral guess).

Subject	N–N	E–E	N–E	E–N	Total
1	1	0	0	1	2
2	1	0	0	0	1
3	2	1	0	0	3
4	0	0	0	0	0
5	1	0	0	0	1
6	0	0	0	0	0
7	0	1	0	0	1
8	0	1	2	2	5
9	0	1	0	0	1
10	0	0	0	0	0
11	1	0	0	0	1
12	6	2	3	4	15
13	1	0	0	0	1
14	1	0	0	0	1
15	0	0	0	1	1
16	0	0	0	1	1
17	0	0	0	0	0
18	0	0	0	1	1
19	1	0	0	0	1
20	0	0	0	0	0
	15	6	5	10	36

The results of the three experiments suggest, therefore, that the nature of the pretask word has little if any effect on the readiness with which subjects will advance hypotheses about the identity of the task word prior to its correct recognition.

DISCUSSION

The present study took as its starting point the assumption that the generalization of perceptual defence is to be explained in terms of set (Forrest *et al.* 1965). The results of the present series of experiments suggest that the type of set involved can be loosely described as a set to see neutral words, rather than a set to perceive pairs of words related in emotional tone. Any more precise description of such a set is elusive, and is not aided by considering 'set' as it is commonly demonstrated in the laboratory. Such demonstrations typically involve facilitated recognition of relatively discrete stimulus classes, of colour or food words, for instance (cf. (Postman & Bruner, 1949; Freeman & Engler, 1955), or of one visual perspective rather than another (Leeper, 1935). The class of 'neutral words' would seem too broad and too ill-defined for its members to benefit from the operation of a comparable set. An alternative would be to postulate a negative set, i.e. a set *not* to see emotional words, which would involve raised rather than lowered thresholds for members of a certain class. But there is no evidence in the literature for the existence of negative set, as opposed to the absence of favourable set, although attempts have been made to explain perceptual defence itself in such terms.

If the nature of a set to see neutral words is unclear, so also is the nature of the state of disruption to which the raised perceptual thresholds are to be ascribed. It is possible that guessing behaviour is inhibited when set is interrupted, but we can find no evidence for any differential readiness to guess after different types of pre-task word. As we have pointed out, however, the total number of pre-recognition hypotheses in our experiments was small, and some differential readiness to guess might still emerge if guessing were to be more strongly encouraged by instructions.

It might alternatively be argued that the disruption of set manifests itself in an inability during the early stages of the recognition process either to perceive or to utilize stimulus information. Inability to perceive stimulus information carries the implication that physical sensitivity, either central or peripheral, is affected in a manner analogous to that proposed by the proponents of the extreme perceptual defence hypothesis (Dixon, 1958; Dixon & Haider, 1961). Inability to utilize stimulus information would, on the other hand, imply the disruption of the process whereby the subject considers evidence derived from part perception in choosing an hypothesis, the procedure described by Woodworth (1947) as 'trial-and-check'.

It would be illuminating to attempt to isolate the effect of partial perception in this situation by eliminating it altogether. Goldiamond & Hawkins (1958) have used a procedure in which the subject is required to 'identify' stimuli which are never presented but which are, in fact, replaced by ambiguous or nonsense material shown repeatedly below recognition threshold. The subject's guesses are scored right or wrong as though the actual stimuli had indeed been presented in a predetermined order. If this technique were employed in the generalization situation and if the generalization effect then appeared, one would have to conclude that any explanation should be couched in terms of independent guessing behaviour rather than by reference to supposed changes in the ability to comprehend the available stimulus information.

The purpose of the present study was to clarify the type of set operative in experiments that purport to show the generalization of perceptual defence. In throwing light on this problem it has revealed two others: the uncertain nature of a set to see neutral words, and the no less obscure characteristics of the disruptive process. These questions require answers with reference to the concept of set itself, answers which it is beyond the scope of this paper to provide.

REFERENCES

DIXON, N. F. (1958). Apparent changes in the visual threshold as a function of subliminal stimulation: a preliminary report. *Quart. J. exp. Psychol.* **10**, 211–9.

DIXON, N. F. & HAIDER, M. (1961). Changes in the visual threshold as a function of subception. *Quart. J. exp. Psychol.* **13**, 229–35.

FORREST, D. W., GORDON, N. & TAYLOR, A. (1965). The generalization of perceptual defence: an interpretation in terms of set. *J. pers. soc. Psychol.* **2**, 137–41.

FREEMAN, J. T. (1954). Set or perceptual defence? *J. exp. Psychol.* **48**, 283–8.

FREEMAN, J. T. & ENGLER, J. (1955). Perceptual recognition thresholds as a function of multiple and single set and frequency of usage of the stimulus material. *Percept. mot. Skills* **5**, 149–54.

GOLDIAMOND, I. & HAWKINS, W. F. (1958). Vexierversuch: the log relationship between word-frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.* **56**, 457–63.

LEEPER, R. (1935). A study of neglected portion of the field of learning—the development of sensory organization. *J. genet. Psychol.* **46**, 41–75.

MCGINNIES, E. & SHERMAN, H. (1952). Generalization of perceptual defence. *J. abnorm. soc. Psychol.* **47**, 81–5.

POSTMAN, L. & BRUNER, J. S. (1949). Multiplicity of set as a determinant of perceptual behaviour. *J. exp. Psychol.* **39**, 369–77.

THORNDIKE, E. L. & LORGE, I. (1944). *The Teacher's Word Book of 30,000 Words.* New York: Teachers' College, Columbia University.

WALTERS, R. H., BANKS, R. K. & RYDER, R. R. (1959). A test of the perceptual defence hypothesis. *J. Pers.* **27**, 47–55.

WOODWORTH, R. S. (1947). Reinforcement of perception. *Amer. J. Psychol.* **60**, 119–24.

(*Manuscript received* 22 *February* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 263–274
Printed in Great Britain

NUMERICAL RELATIONS, SIMILARITY, AND SHORT-TERM RECOGNITION MEMORY FOR PAIRS OF DIGITS

BY WAYNE A. WICKELGREN
Massachusetts Institute of Technology

Recognition memory for pairs of digits after a 6 sec interference task was superior for pairs containing a zero or one and for pairs consisting of two identical digits than for 'ordinary' pairs. Pairs consisting of digits in forward or backward sequence (e.g. 56 or 43) and pairs where one digit is a multiple of the other were remembered slightly better than 'ordinary' pairs. False recognition rates were highest for test pairs that had the highest degrees of 'identical elements' similarity to the presented pair (two identical digits in reversed positions or one identical digit in the same position as in the presented pair). The results are discussed in terms of an associative theory of short-term recognition memory.

Most people feel intuitively that pairs of digits such as 44, 67, 32, 84 are easier to remember than pairs of digits such as 37, 85, 92, 83. The first set consists of pairs whose digits have some familiar numerical relationship to each other—identity (44), forward numerical order (67), backward numerical order (32), multiple (84). The second set consists of pairs whose digits have no universally familiar relationship.

There are some studies on short-term recall of lists of six to ten items containing runs of identical items (44, 999, etc.) by Ranschburg (1902), Kleinknecht (1906), Turley (1906), Obonai & Tatsuno (1954), Tatuno (1961), and Wickelgren (1965). These studies suggest that a run of two identical digits is easier to remember than two different digits, but this result is affected by the position of the run in the list and the speed of presentation. Furthermore, these experiments all used lists much longer than two. There are apparently no studies on short-term memory for pairs whose digits have numerical relations other than identity.

Although one's first intuition is that pairs with familiar relations ought to be easier to remember than pairs without familiar relations, it is not easy to justify this conclusion theoretically. It is true that, if the relation is recognized, one can recode a pair of digits into one digit and a relation but, in English, this seems always to require as many or more 'words' than merely storing the two digits without recoding. It might also be claimed that the activity of counting forwards and backwards by ones, twos, threes, and so on, has differentially strengthened the associations between certain digits and not others. However, much the greater part of our experience with digits does not consist of counting, and any differences in the frequency of experiencing different pairs in sequence must be very small in relation to the total frequency of exposure to each pair. Existing data suggest that verbal learning is a negatively accelerated function of frequency of past experience (Underwood & Schulz, 1960), which implies that large differences in absolute frequency of past experience make little difference, once the absolute frequency of past experience is large. Even the most infrequent pair of digits has been experienced a gigantic number of times by the average college student. Thus it is by no means obvious that pairs with familiar relations will be easier to remember than pairs without familiar relations.

An experiment on letter–number paired associate learning by Williams (1961) suggested that some pairs of digits may be easier to remember than others, but the categorization of digit pairs was not on the basis of numerical relations. Williams found evidence that numbers from 1 to 15 were easier to remember than numbers from 16 to 50. Since the numbers from 1 to 9 were presented as single digits rather than as pairs, 01,..., 09, it is not clear whether this result reflects anything beyond the greater ease of remembering one digit than two digits. However, the fact that Williams included the pairs 10–15 in the 'easy' group argues that perhaps pairs beginning with 0 and 1 are easier than pairs containing no 0 or 1. Maybe pairs ending in 0 or 1 are also easier than pairs containing no 0 or 1. Again, it is not obvious that this should be so on the basis of either recoding or past experience, for reasons completely analogous to those stated in the previous paragraph.

The first objective of the present study was to determine if some types of digit pairs are easier to remember than other types of digit pairs, and, if so, which types. The second objective was to determine if a recognition memory test pair that is 'similar' to the originally presented pair in some way will produce a higher frequency of false recognition than a test pair that is less similar. There are many ways in which the test pair might be similar to the pair originally presented, such as, for instance, possessing one or two common digits in the same or different positions, or possessing the same type of numerical relation between the digits in the pair. To the extent that the memory trace for a pair of digits is composed of at least semi-independent traces for its component digits and their relations, it should be possible to demonstrate a higher frequency of false recognition for test pairs that have some features in common with the presented pair. If the memory trace for a pair, such as 95, is a completely additive combination of separate traces for 9 and for 5 and for 'decreasing order' or a completely additive combination of separate traces for 9 in the first position and 5 in the second position, then the false recognition rate for highly similar test pairs ought to be an appreciable fraction of the correct recognition rate. If the memory trace for a pair is highly super-additive to the point of being a completely contingent association—that is, the trace for '95' is not the sum of any set of traces for the component properties of '95', then there should be no difference in false recognition frequency for more or less similar test pairs.

METHOD

Experiment I

Procedure

On each trial a ready signal was followed after 1 sec by a pair of digits to be copied during the second in which the pair was presented and the 0·5 sec interval that followed (filled with a tone). Immediately after the tone, twelve randomly selected interference digits were presented at the rate of 0·5 sec/digit; subjects were to copy these as they were being presented. Immediately after the interference task, a tone sounded for about 0·5 sec, followed by a recognition test pair of digits, which was also to be copied. Subjects were then given 10 sec in which to decide whether the test pair was identical with the presented pair or not and to indicate their confidence in this decision on a five-point rating scale. The subjects' label for these decisions was 'yes' or 'no', respectively. Naturally, subjects were instructed to cover what they had copied for the originally presented pair immediately after copying the presented pair. The purpose of requiring subjects to copy the originally presented pair was to ensure that subjects paid attention to the pair to the point of correct recognition and to limit the time available for rehearsal of the pair by forcing subjects to make responses during the presentation of the pair. Each trial required a total of about 20 sec. The entire experiment was recorded on tape and administered to groups of about thirteen subjects each. The experiment lasted about 1 hr.

Design

There were thirty-one conditions in the experiment, differing in the nature of the presented pair and in the similarity of the test pair to the presented pair. The order of presenting the conditions was randomized in blocks of thirty-one trials, and there were five blocks in the experiment for a total of 155 trials. The complete list of the thirty-one conditions is given in Table 1. Subjects were thirty-eight M.I.T. undergraduates who participated in the experiment to fulfil a requirement of their psychology courses.

Table 1. *Conditions in Experiment I*

Condition	Type presented pair	Type of test pair	Relation of test pair to presented pair
$(XX;\ XX)$	*I*	*I*	Identical
$(XX;\ YY)$	*I*	*I*	No common items
$(XX;\ XY)$	*I*	*NOR*	First item identical. Second different
$(XX;\ YZ)$	*I*	*NOR*	No common items
$(mX, nX;\ mX, nX)$	*M*	*M*	Identical
$(mX, nX;\ nX, mX)$	*M*	*M*	Identical items in reverse order
$(mX, nX;\ jY, kY)$	*M*	*M*	No common items
$(mX, nX;\ mX, Y)$	*M*	*NOR*	First item identical
$(mX, nX;\ Y, Z)$	*M*	*NOR*	No common items
$(X, X+1;\ X, X+1)$	*FO*	*FO*	Identical
$(X, X+1;\ X+1, X)$	*FO*	*BO*	Identical items in reverse order
$(X, X+1;\ Y, Y+1)$	*FO*	*FO*	No common items
$(X, X+1;\ X, Y)$	*FO*	*NOR*	First item identical
$(X, X+1;\ Y, Z)$	*FO*	*NOR*	No common items
$(X, X-1;\ X, X-1)$	*BO*	*BO*	Identical
$(X, X-1;\ X-1, X)$	*BO*	*FO*	Identical items in reverse order
$(X, X-1;\ Y, Y-1)$	*BO*	*BO*	No common items
$(X, X-1;\ X, Y)$	*BO*	*NOR*	First item identical
$(X, X-1;\ Y, Z)$	*BO*	*NOR*	No common items
$(XY;\ XY)$	*NOR*	*NOR*	Identical
$(XY;\ YX)$	*NOR*	*NOR*	Identical items in reverse order
$(XY;\ XZ: S)$	*NOR*	*NOR*	First item identical: same order
$(XY;\ XZ: D)$	*NOR*	*NOR*	First item identical: different order
$(XY;\ ZX: S)$	*NOR*	*NOR*	Second item of test is first item of presented pair: same order
$(XY;\ ZX: D)$	*NOR*	*NOR*	Second item of test is first item of presented pair: different order
$(XY;\ ZY: S)$	*NOR*	*NOR*	Second item identical: same order
$(XY;\ ZY: D)$	*NOR*	*NOR*	Second item identical: different order
$(XY;\ YZ: S)$	*NOR*	*NOR*	First item of test is second item of presented pair: same order
$(XY;\ YZ: D)$	*NOR*	*NOR*	First item of test is second item of presented pair: different order
$(XY;\ WZ: S)$	*NOR*	*NOR*	No common items: same order
$(XY;\ WZ: D)$	*NOR*	*NOR*	No common items: different order

Note. *I* stands for a pair of identical digits such as 44. *M* stands for two digits that are multiples of a digit $\geqslant 2$, such as 68, 24, or 39. *FO* stands for two digits in forward numerical order, such as 56. *BO* stands for two digits in backward numerical order, such as 87. *NOR* stands for two digits that have 'no common relation', which means the pair is not *I*, *M*, *FO*, or *BO*. $(XY:\ WZ)$ are in the 'same order' if and only if ($X > Y$ implies $W > Z$) and ($X < Y$ implies $W < Z$). 'Different order' is the reverse of 'same order'.

Experiment II

Procedure

The procedure differed from the first experiment in only one way. In the second experiment, subjects were required to copy four proactive-interference (PI) digits before being presented with the pair to be remembered later. The four PI digits were presented at the rate of 0·5 sec/digit, and were followed immediately by the presented pair, which was copied as in the first experiment. The purpose of including the PI digits in this experiment was to increase the error rate, which is low for recognition memory of a single pair of digits.

Table 2. *Conditions in Experiment II*

Condition	Type of presented pair	Type of test pair	Relation of test pair to presented pair
(*XX; XX*)	*I*	*I*	Identical
(*XX; YZ*)	*I*	*NOR*	No common items
(*X, nX; X, nX*)	*M*	*M*	Identical
(*X, nX; YZ: S*)	*M*	*NOR*	No common items: same order
(*X, X+1; X, X+1*)	*FO*	*FO*	Identical
(*X, X+1; YZ : S*)	*FO*	*NOR*	No common items: same order
(*X, X−1; X, X−1*)	*BO*	*BO*	Identical
(*X, X−1; YZ: S*)	*BO*	*NOR*	No common items: same order
(*XY; XY*)	*NOR*	*NOR*	Identical
(*XY; ZZ*)	*NOR*	*I*	No common items
(*XY; Z, nZ: S*)	*NOR*	*M*	No common items: same order
(*XY; Z, Z+1: S*)	*NOR*	*FO*	No common items: same order
(*XY; Z, Z−1: S*)	*NOR*	*BO*	No common items: same order
(*XY; XZ: S*)	*NOR*	*NOR*	First item identical: same order
(*XY; ZY: S*)	*NOR*	*NOR*	Second item identical: same order
(*XY; WZ: S*)	*NOR*	*NOR*	No common items: same order
(*XY; WZ: D*)	*NOR*	*NOR*	No common items: different order
(*XY; 0Z: S*)	*NOR*	*0X*	No common items: same order
(*XY: Z0: S*)	*NOR*	*X0*	No common items: same order
(*XY; 1Z: S*)	*NOR*	*1X*	No common items: same order
(*XY; Z1: S*)	*NOR*	*X1*	No common items: same order
(*0X; 0X*)	*0X*	*0X*	Identical
(*0X; YZ: S*)	*0X*	*NOR*	No common items: same order
(*X0; X0*)	*X0*	*X0*	Identical
(*X0; YZ: S*)	*X0*	*NOR*	No common items: same order
(*1X; 1X*)	*1X*	*1X*	Identical
(*1X; YZ: S*)	*1X*	*NOR*	No common items: same order
(*X1; X1*)	*X1*	*X1*	Identical
(*X1; YZ: S*)	*X1*	*NOR*	No common items: same order

Note. *I*, *M*, *FO*, *BO*, and *NOR* have the same meanings as in Table 1, except that *W*, *X*, *Y*, and *Z* cannot be 0 or 1 and *M* is restricted to multiples of 2 or 4 where one member of the pair is also a multiple of the other, such as 24, 42, 28, 82, 48, 84. (*X*, *nX*) is only a convenient abbreviation for (*X*, *nX* or *nX*, *X*). *0X* means an *NOR* pair in which zero is the first element. *X0*, *1X*, and *X1* are completely analogous.

Design

There were thirty-one conditions in the second experiment also, but most of the conditions were somewhat different from those in the first experiment. The order of presenting the conditions was randomized in blocks of thirty-one trials, and there were again five blocks of trials, as in Expt I. Subjects were forty-four M.I.T. undergraduates taking psychology courses. The complete list of conditions is given in Table 2.

RESULTS

On each trial a subject must choose one of ten decision-confidence pairs. Let $i = 1, \ldots, 5, 6, \ldots, 10$ stand for 'yes' with confidence '5' (greatest confidence),..., 'yes' with confidence '1' (least confidence), 'no' with confidence '1',..., 'no' with confidence '5': $f_i(x)$ may then represent the total frequency (over all blocks and all subjects) with which response i occurred in condition x;

$$r_i(x) = f_i(x) / \sum_{i=1}^{10} f_i(x)$$

may represent the relative frequency with which response i occurred in condition x; and

$$R_i(x) = \sum_{j=1}^{i} r_i(x)$$

may represent the cumulative relative frequency with which responses 1 to i occurred in condition x. The cumulative relative frequency functions obtained for each condition in Expts I and II are given in Tables 3 and 4, respectively.

Table 3. *Values (%) of* $R_i(x)$ *in Experiment I*

$R_i(x)$ *in Experiment I*

Condition (x)	$Y5$	$Y4$	$Y3$	$Y2$	$Y1$	$N1$	$N2$	$N3$	$N4$	$N5$	N
$(XX; XX)$	82·6	88·4	90·5	90·5	90·5	91·0	92·1	93·7	94·8	100	190
$(XX; YY)$	0·0	1·1	2·7	3·2	3·2	4·3	5·4	8·6	14·4	100	190
$(XX; XY)$	1·1	1·6	2·7	2·7	2·7	3·2	4·3	6·9	13·2	100	190
$(XX; YZ)$	0·0	0·5	0·5	0·5	0·5	0·5	1·0	5·2	14·1	100	190
$(mX, nX; mX, nX)$	50·5	60·0	68·4	70·5	72·6	75·2	77·3	83·1	88·9	100	190
$(mX, nX; nX, mX)$	0·5	2·1	2·6	3·1	3·1	3·6	5·2	13·1	24·2	100	190
$(mX, nX; jY, kY)$	0·5	0·5	0·5	1·6	1·6	2·1	2·6	9·4	18·3	100	190
$(mX, nX; mX, Y)$	2·1	4·7	6·3	6·3	6·8	7·9	10·5	14·7	28·9	100	190
$(mX, nX; Y, Z)$	0·5	1·6	1·6	2·1	2·6	2·6	3·1	6·3	16·8	100	190
$(X, X+1; X, X+1)$	57·9	69·5	75·8	77·4	77·9	79·5	80·6	86·9	92·2	100	190
$(X, X+1; X+1, X)$	1·1	4·3	7·5	7·5	8·6	8·6	8·6	17·0	29·1	100	189
$(X, X+1; Y, Y+1)$	0·5	0·5	0·5	1·0	1·0	1·0	1·5	4·1	13·6	100	188
$(X, X+1; X, Y)$	0·0	1·1	1·1	1·1	1·6	2·7	4·3	9·6	21·7	100	190
$(X, X+1; Y, Z)$	0·5	1·6	3·7	4·2	4·7	5·2	6·8	10·5	23·7	100	190
$(X, X-1; X, X-1)$	61·6	73·2	80·0	82·6	83·1	84·7	86·3	91·0	94·2	100	189
$(X, X-1; X-1, X)$	1·1	2·2	3·8	4·9	5·4	7·5	9·1	19·1	33·8	100	189
$(X, X-1; Y, Y-1)$	1·1	1·1	2·2	2·2	2·2	4·3	4·8	12·7	24·3	100	190
$(X, X-1; X, Y)$	0·0	0·0	0·5	0·5	0·5	1·0	1·5	4·1	17·8	100	190
$(X, X-1; Y, Z)$	0·0	0·0	0·5	1·0	1·5	2·6	3·1	8·4	20·0	100	190
$(XY; XY)$	58·4	73·1	78·9	79·4	80·5	81·6	84·8	90·6	94·3	100	190
$(XY; YX)$	0·0	1·6	3·2	3·2	3·7	3·7	5·3	16·9	30·6	100	190
$(XY; XZ: S)$	2·6	4·7	6·3	6·8	6·8	7·3	8·9	12·6	24·2	100	190
$(XY; XZ: D)$	1·1	1·6	2·1	3·2	3·2	3·2	4·8	12·2	24·3	100	190
$(XY; ZX: S)$	0·0	0·5	1·0	1·0	1·0	1·5	2·0	10·9	21·4	100	190
$(XY; ZX: D)$	0·5	0 5	2·1	2·1	2·6	5·2	6·8	9·4	19·4	100	188
$(XY; ZY: S)$	3·2	6·4	9·0	11·1	11·1	12·2	14·3	22·2	34·3	100	190
$(XY; ZY: D)$	1·1	1·6	3·7	4·8	6·4	8·5	10·1	12·7	28·5	100	189
$(XY; YZ: S)$	1·1	1·1	2·2	2·2	2·2	3·3	4·4	8·1	20·7	100	190
$(XY; YZ: D)$	0·0	0·0	0·5	1·0	1·5	2·0	3·6	5·7	19·9	100	190
$(XY; WZ: S)$	0·0	2·1	3·2	3·2	3·7	5·8	6·3	11·6	20·5	100	190
$(XY; WZ: D)$	0·5	0·5	1·0	1·5	1·5	2·0	4·6	9·9	24·6	100	190

Similarity of test to presented pair

Same versus different numerical order. There are five pairs of conditions in Expt I [(*XY; XZ: S*) versus (*XY; XZ: D*), (*XY; ZX: S*) versus (*XY; ZX: D*), (*XY; ZY: S*) versus (*XY; ZY: D*), (*XY; YZ: S*) versus (*XY; YZ: D*), and (*XY; WZ: S*) versus (*XY; WZ: D*)] and one pair of conditions in Expt II [(*XY; WZ: S*) versus (*XY; WZ: D*)] that were designed to determine whether the numerical order of the digits in a pair is an important cue in recognition memory. In the present context, 'numerical order' simply refers to whether the first digit is greater than the second digit, e.g. 95, or vice versa, e.g. 59. On the average, there is a higher false recognition rate in the conditions where the test pair and the presented pair have the same order than in the conditions where they have a different order, other things being equal. However, the effect is far from significant if χ^2 tests are used to compare 'yes' and 'no' responses in each condition, and there is a non-significant, opposite effect in some cases. By itself, numerical order is not an important cue in short-term recognition memory of digit pairs. This means either that numerical order is never an important cue or that it only contributes to the probability of recognizing a pair when there are other functioning cues. Some support for the latter alternative can be

Table 4. *Values* (%) *of* $R_t(x)$ *in Experiment II*

$R_t(x)$ *in Experiment II*

Condition (*x*)	*Y*5	*Y*4	*Y*3	*Y*2	*Y*1	*N*1	*N*2	*N*3	*N*4	*N*5	*N*
(*XX; XX*)	70·4	81·2	83·9	85·6	87·4	87·9	90·1	91·5	91·9	100	223
(*XX; YZ*)	0·5	0·5	3·3	3·8	4·2	8·5	12·2	21·1	30·0	100	213
(*X, nX; X, nX*)	35·3	47·7	60·8	64·4	67·6	72·7	77·3	86·5	91·7	100	436
(*X, nX; YZ: S*)	3·6	5·8	12·9	17·8	20·0	26·2	32·9	43·1	50·2	100	225
(*X, X+1; X, X+1*)	46·6	59·3	67·9	70·6	74·2	80·1	81·4	87·8	91·8	100	221
(*X, X+1; YZ: S*)	2·4	6·2	8·6	10·9	12·9	19·0	23·3	38·1	51·4	100	210
(*X, X−1; X, X−1*)	36·9	46·2	57·8	60·4	64·4	70·7	76·0	83·6	91·1	100	225
(*X, X−1; YZ: S*)	1·9	3·3	7·1	10·4	11·9	16·6	19·4	31·3	46·4	100	211
(*XY; XY*)	36·4	48·6	61·0	67·2	72·0	73·3	82·8	89·0	95·0	100	418
(*XY; ZZ*)	0·0	0·5	1·4	2·4	2 4	4·7	7·6	18·7	31·6	100	212
(*XY; Z, nZ: S*)	0·9	1·4	4·6	5·1	5 5	11·0	17·0	33·9	49·1	100	218
(*XY; Z, Z+1: S*)	0·9	2·3	3·2	4·2	5·5	12·4	18·4	33·2	45·2	100	217
(*XY; Z, Z−1: S*)	0·5	1·5	2·9	4·4	5·8	12·1	18·8	30·0	42·5	100	207
(*XY; XZ: S*)	2·4	3·8	9·4	12·2	13·6	16·4	24·4	38·5	53·0	100	213
(*XY; ZY: S*)	2·4	4·7	10·9	13·2	15·6	21·7	28·3	40·1	54·2	100	212
(*XY; WZ: S*)	1·3	3·1	8·9	10·2	10·7	16·4	23·6	35·5	47·6	100	225
(*XY; WZ: D*)	0·5	1·8	7·6	10·3	13·1	18·0	24·3	38·3	51·8	100	222
(*XY; 0Z: S*)	3·1	4·4	6·2	8·4	10·2	15·5	21·2	28·8	41·6	100	226
(*XY; Z0: S*)	0·5	0·9	0·9	2·3	2·3	5·1	12·6	26·6	37·8	100	214
(*XY; 1Z: S*)	0·0	0·5	1·4	1·9	2·4	5·2	12·8	23·7	38·4	100	211
(*XY; Z1: S*)	0·0	0·0	1·9	2·8	2·8	7·5	14·6	27·2	41·3	100	213
(*0X; 0X*)	50·9	64·1	73·6	78·8	81·6	84·9	85·4	92·4	94·8	100	212
(*0X; YZ: S*)	0·0	0·0	0·5	0·5	0·9	3·7	11·7	24·3	37·8	100	214
(*X0; X0*)	52·5	69·5	76·2	80·3	81·6	83·9	85·2	90·1	93·7	100	223
(*X0; YZ: S*)	0·9	1·9	4·2	5·2	6·1	8·5	11·7	22·5	34·7	100	213
(*1X; 1X*)	47·6	56·1	64·6	70·7	73·1	75·5	79·2	83·9	89·6	100	212
(*1X; YZ: S*)	0·5	1·4	3·2	4·1	4·6	9·5	16·8	30·0	46·8	100	220
(*X1; X1*)	32·3	42·4	53·9	61·3	63·6	67·3	73·3	81·6	89·9	100	217
(*X1; YZ: S*)	0·5	1·9	6·2	9·1	10·5	15·2	20·0	33·3	46·7	100	210

derived from the fact that the largest positive effects of numerical order were obtained in the conditions where the test pair had the highest degree of similarity to the presented pair (first or second digits identical).

Same versus different items and position. There are four pairs of conditions in Expt I [$(mX, nX; nX, mX)$ versus $(mX, nX; Y, Z)$, $(X, X+1; X+1, X)$ versus $(X, X+1; Y, Z)$, $(X, X-1; X-1, X)$ versus $(X, X-1: Y, Z)$, and $(XY; YX)$ versus $(XY; WZ: D)$] that were designed to determine whether the false recognition rate is higher for test pairs that consist of the same two digits as those in the presented pair, but in the wrong positions. Each of the four comparisons is in the direction of greater false recognition rate for pairs having the two presented digits in reverse positions than for pairs having no digits in common with the presented pair, and the overall comparison is significant ($\chi^2 = 5{\cdot}76$, D.F. 1, $P < 0{\cdot}02$).

Comparison of conditions $(XY; XZ: S)$, $(XY; XZ: D)$, $(XY; ZY: S)$, and $XY; ZY: D)$ with $(XY; WZ: S)$ and $(XY; WZ: D)$ in Expt I provides a test of the hypothesis that the false recognition rate is greater for test pairs having one digit in common with the presented pair in the same position than for pairs having nothing in common with the presented pair. The results support the hypothesis and the effect is significant ($\chi^2 = 7{\cdot}96$, D.F. 1, $P < 0{\cdot}01$). In addition, the effect on false recognition rate appears to be greater for the second item than for the first item of the pair. The analogous comparison of conditions $(XY; XZ: S)$, $(XY; ZY: S)$, and $(XY; WZ: S)$ in Expt II supports the same conclusions, but is not significant. The three comparisons of 'first items identical' with 'no common items' for *M*, *FO*, and *BO* presented pairs in Expt I are non-significant and inconsistent. However, for *NCR* pairs it is clear that false recognition rate is greater for test pairs with either the first or second item identical with the presented pair.

When the test pair and the presented pair contain one common digit, but in the wrong position, there is no significant effect on false recognition rate.

Nature of the digits (0, 1, 2–9)

There are three, theoretically distinct, ways of comparing recognition memory for pairs of the form *0X*, *X0*, *1X*, *X1*, and *XY*. One can compare false recognition rates for test pairs of the form *WZ* following each of the different types of presented pairs. One can compare false recognition rates for different types of test pairs following presented pairs of the form *WZ*. Finally, one can compare correct recognition rates for each of the pair types. If pairs with 0 and 1 in them are easier to remember, then it should be easier to reject incorrect test pairs of the form *WZ* when the presented pair had a 0 or 1 in it, and it should be easier to reject an incorrect test pair with a 0 or 1 in it after a presented pair of the form *WZ*. Finally, it should be easier to recognize the presented pair correctly.

In order to evaluate the strength of the memory trace for pairs of different types, it is necessary to be sure that the 'non-memory' response biases are the same for test pairs of different types. This is obviously no problem in the first of the above comparisons because all the test pairs are of the same type. However, in the last two comparisons, there may be some non-memory differences in the tendency to say 'yes' to test pairs of different types. The analogous problem occurs in signal detect-ability experiments, where the solution to the problem has been to compute the 'receiver operating characteristic' (ROC curve) for each condition. Applied to an experiment in recognition memory this function is more appropriately called the 'memory operating characteristic' (MOC curve). The MOC curve is a plot of $R_i(x)$, the 'correct recognition' rate for some condition x in which the probe pair is identical with the presented pair, against $R_i(y)$, the 'false recognition' rate for some control condition y in which the probe pair is of the same type as in condition x, but not identical with the presented pair. Since the MOC curve is a plot of two cumulative probability functions, one against the other, the curve originates at (0, 0) and ends at (1, 1). If one type of digit pair is remembered better than another type of digit pair, then the MOC curve for the first type of digit pair will lie above the MOC curve for the second type of digit pair. This means that, for all values of the false recognition rate, the correct recognition rate for the first type of pair is greater than the correct recognition rate for the second type of pair.

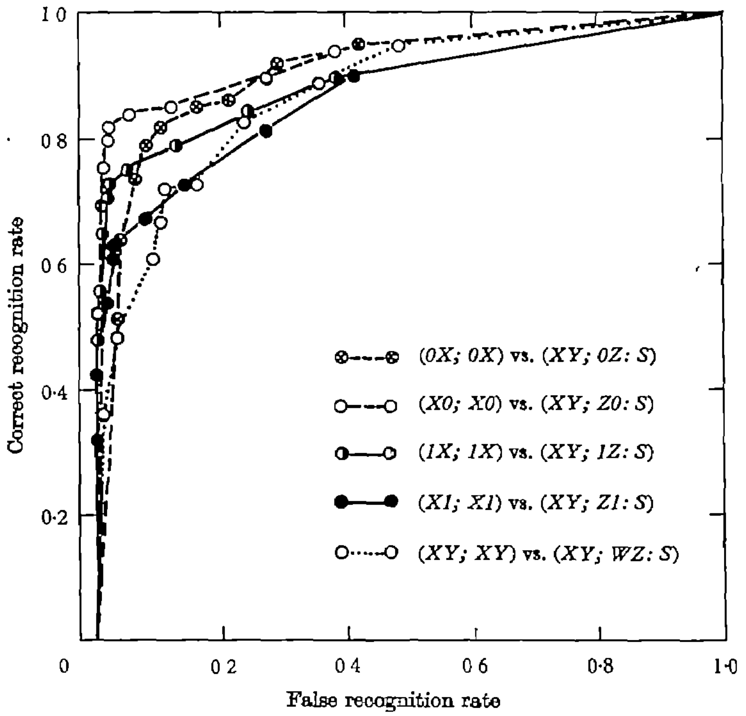


Fig. 1. 'Memory operating characteristic' (MOC) curves for *0X*, *X0*, *1X*, *X1*, and *XY* pairs in Expt. II.

False recognition rates for presented pairs containing 0 or 1. Conditions (*0X*; *YZ*: *S*), (*X0*; *YZ*: *S*), (*1X*; *YZ*: *S*), (*X1*; *YZ*: *S*) and (*XY*; *WZ*: *S*) in Expt II have test pairs of the same type and differ only in the nature of the presented pair. Since (*XY*; *WZ*: *D*) does not differ from (*XY*; *WZ*: *S*), these two conditions are combined for all subsequent statistical analyses to increase the number of cases. False recognition rates for the above five types of presented pairs fall into three categories. Presented pairs beginning with zero (*0X*) produce a significantly lower false recognition rate than presented pairs ending with zero (*X0*) and beginning with one (*1X*), which in turn have a lower false recognition rate than pairs ending in one (*X1*) or having no zero or one (*XY*) ($\chi^2 = 6{\cdot}25$, D.F. 1, $P < 0{\cdot}02$ and $\chi^2 = 11{\cdot}12$, D.F. 1, $P < 0{\cdot}001$, respectively.

False recognition rates for test pairs containing 0 or 1. Conditions (*XY; 0Z: S*), (*XY; Z0: S*), (*XY; 1Z: S*), (*XY; Z1: S*) and (*XY; WZ: S*) in Expt II have presented pairs of the same type and differ only in the nature of the test pair. False recognition rates are substantially lower for test pairs containing a 0 or 1 ($\chi^2 = 23{\cdot}23$, D.F. 1, $P < 0{\cdot}001$), but test pairs of the form *0X* appear to have a higher false recognition rate than pairs of the form *X0*, *1X*, or *X1*.

Correct recognition rates for pairs containing 0 or 1. Comparison of the correct recognition rate in conditions (*0X; 0X*), (*X0*; *X0*), (*1X; 1X*), and (*X1; X1*) with (*XY; XY*) in Expt II is biased against the conditions with 0's or 1's in them because of their lower false recognition rate. In spite of this, a Kolmogorov–Smirnov test on the cumulative recognition functions shows that the correct recognition rate is higher in (*0X; 0X*), (*X0; X0*), and (*1X; 1X*) than in (*XY; XY*) [$D = 15{\cdot}6\%$, $P < 0{\cdot}01$; $D = 20{\cdot}9\%$, $P < 0{\cdot}001$; $D = 11{\cdot}3\%$, $P < 0{\cdot}06$, respectively]. The correct recognition rate in (*X1; X1*) is not significantly different from the rate in (*XY; XY*). The MOC curves for each of the five types of pairs is presented in Fig. 1. MOC curves control for differences in response biases by plotting correct recognition rate as a function of false recognition rate for each type of test pair, (*0X; 0X*) versus (*XY; 0Z: S*), (*X0; X0*) versus (*XY; Z0: S*), (*1X; 1X*) versus (*XY; 1X: S*), (*X1; X1*) versus (*XY; X1: S*), and (*XY; XY*) versus (*XY; WZ: S*). The MOC curves indicate that pairs with a zero in the first position, a zero in the second position, or a one in the first position are easier to remember than pairs with a one in the second position or pairs with no zero or one.

Nature of the relation between the digits (I, M, FO, BO, NCR)

False recognition rates for different types of presented pairs. Conditions (*XX; YZ*), (*X, nX; YZ: S*), (*X, X+1; YZ: S*), (*X, X−1; YZ: S*) and (*XY; WZ: S*) in Expt II have test pairs of the same type and differ only in the nature of the relationship of the digits in the presented pair. The false recognition rate is lower for *I* and higher for *M* presented pairs than for *NCR* presented pairs ($\chi^2 = 8{\cdot}99$, D.F. 1, $P < 0{\cdot}01$ and $\chi^2 = 7{\cdot}33$, D.F. 1, $P < 0{\cdot}01$, respectively). *FO* and *BO* presented pairs did not differ significantly from *NCR* presented pairs. The analogous comparisons in Expt II showed slightly lower false recognition rates for *I*, *M*, and *BO* pairs than for *NCR* and *FO* pairs, but no differences were significant.

False recognition rates for different types of test pairs. Conditions (*XY; ZZ*), (*XY; Z, nZ: S*), (*XY; Z, Z+1: S*), (*XY; Z, Z−1: S*), and (*XY; WZ: S*) in Expt II have the same type of presented pair and differ only in the nature of the relationship of the digits in the test pair. The false recognition rate is lower for *I*, *M*, *FO*, and *BO* test pairs than for *NCR* test pairs ($\chi^2 = 15{\cdot}00$, D.F. 1, $P < 0{\cdot}001$; $\chi^2 = 5{\cdot}98$, D.F. 1, $P < 0{\cdot}02$; $\chi^2 = 5{\cdot}92$, D.F. 1, $P < 0{\cdot}02$; $\chi^2 = 5{\cdot}15$, D.F. 1, $P < 0{\cdot}02$, respectively).

Correct recognition rates for I, M, FO, BO, and NCR pairs. Comparisons of the correct recognition rates in conditions (*XX; XX*), (*X, nX; X, nX*), (*X, X+1; X, X+1*) and (*X, X−1; X, X−1*) with (*XY; XY*) is biased against the *I*, *M*, *FO* and *BO* conditions because of their lower false recognition rate. In spite of this, a Kolmogorov–Smirnov test on the cumulative recognition functions shows that the correct recognition rate is significantly greater for *I* and *FO* pairs than *NCR* pairs ($D = 34\%$, $P < 0{\cdot}001$ and $D = 13{\cdot}2\%$, $P < 0{\cdot}02$, respectively). *M* and *BO* pairs are not significantly different from *NCR* pairs. The analogous comparisons in Expt I show a significantly higher correct recognition rate for *I* pairs than for *NCR* pairs ($D = 24{\cdot}2\%$, $P < 0{\cdot}001$) and no significant differences among *M*, *FO*, *BO*, *NCR* pairs.

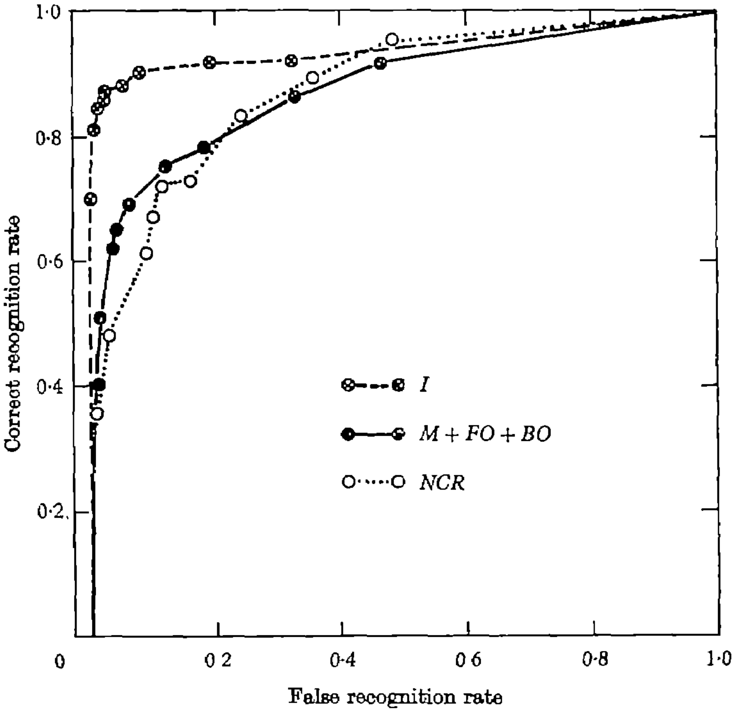


Fig. 2. 'Memory operating characteristic' (MOC) curves for *I*, *M*, *FO*, *BO*, and *NCR* pairs in Expt II.

The MOC curves for the pairs with different types of relationships between their digits are shown in Fig. 2. *M*, *FO*, and *BO* pairs have been averaged into one curve to make the figure clearer. Pairs consisting of two identical digits are remembered much better than the other types of pairs, and there appears to be some superiority for *M*, *FO*, and *BO* pairs over *NCR* pairs, but the latter difference is much smaller than the former.

DISCUSSION

Unless one uses a paired associate procedure where every response item has its own special stimulus item, one experiences a peculiar kind of frustration in specifying the stimulus for recall of the first item of a serial list. The frustration is quite acute if there is only one item in the list, because then the problem is difficult to ignore. Within an associative theory of memory the solution to the problem may be as follows. Before and during presentation of an item, a well-instructed and well-motivated subject is thinking something like '—— is the presented item' and therefore, by simple contiguity, associations will be formed in both directions between the internal representative of the presented item and the internal representative of

'—— is the presented item'. At the time of recall a subject is able to activate the internal representative of '—— is the presented item'; and, if the association from '—— is the presented item' to the representative of the presented item is strong enough, then the presented item is recalled.

Recognition memory simply tests the strength of the reverse association, from the representative of the test item to the representative of '—— is the presented item'. If the association is strong enough to activate the representative of '—— is the presented item', then the subject says 'yes'. The exact nature of the internally generated concept, which has been symbolized by '—— is the presented item', is irrelevant to the present discussion. It makes no difference what the exact words are or even whether the concept is verbalizable at all. The conscious or unconscious nature of the concept activation process is also irrelevant, as are individual differences.

In the present study the presented 'item' is a pair of digits. During presentation of a pair, an association is formed between the representative of the presented pair and the representative of '—— is the presented pair'. When the test pair is identical to the presented pair, then the strength of this association is being tested. When the test pair is different from the presented pair, then what is being tested is the strength of the association between the representative of the test pair and the representative of '—— is the presented pair'.

If the internal representative of 74 has no more in common with the internal representatives of 47, 73, or 94 than it has in common with the internal representatives of 58, 92, 35, etc., then we should expect the false recognition rates to be identical for incorrect test pairs with different degrees of similarity to the presented pair. This is not the case. Test pairs with the higher degrees of 'identical elements' similarity (two identical digits in reversed positions or one identical item in the identical position) have significantly higher false recognition rates than test pairs with lower degrees of similarity. This suggests that the internal representative of 74 is composed of the sequential activation of the internal representatives of 7 and 4. The representative of 7 provides input to the representative of '—— is the presented pair'; the representative of '7 in the first position' provides even more input to the representative of '—— is the presented pair'.

The results of the present study clearly indicate that the probability of activating '—— is the presented pair' in response to the correct pair is not a simple multiplicative function of the probability of activating '—— is the presented pair' in response to each element of the pair. The evidence for this assertion is that the ratio $R_i(XY;\ XY)/R_i(XY;\ XZ{:}S)$ is much greater than the ratio $R_i(XY;\ XZ{:}S)/R_i(XY;\ WZ{:}S)$ for all values of i. Other ratio tests also fail to show the equality predicted by the simple multiplicative probability model. The failure of the simple multiplicative model does not imply that the representative of the whole is an indivisible unit, unrelated to the representatives of its parts. In fact, this latter hypothesis is rather unlikely in view of the relationship between similarity and false recognition. What seems to be indicated by the data is that the representative of each element of the pair provides input to the representative of '—— is the presented pair', but there is a non-linear relationship between the amount of input and the probability of activation. One such non-linear relation would be a threshold function.

The only type of similarity demonstrated to affect false recognition rate in the present experiments was identical elements similarity. The only type of relational similarity investigated in a systematic manner was the numerical order of the two digits in the pair, and numerical order had only a small and insignificant effect on false recognition rate. It seems likely that, if there is an effect of any type of relational similarity, it is small enough to require a much larger number of trials or subjects in order to demonstrate it than were used in the present experiments.

Pairs of the form *X1* appear to be remembered slightly better than pairs of the form *XY*; pairs of the form *0X*, *X0*, *1X*, and *XX* are remembered much better than *XY* pairs. Pairs with numerical relations more complex than identity (such as multiples, forward sequences, and backward sequences) are remembered a little better than *XY* pairs with no such numerical relations between their digits, but the difference is quite small. The present study was designed to demonstrate the existence of such differences and their approximate magnitude. Explanation of the differences seems difficult, but one can hazard the guess that subjects classify pairs into types. This means that a presented pair activates other concepts in associative memory that become associated to '—— is the presented pair' and thus provide additional cues for correct recognition. It seems likely that the defining concepts for 'special' pairs are simpler, for example, in number of 'words'. 'Ordinary' pairs are very likely defined negatively, that is, 'not two identical items', 'not containing a 0', etc. If there is only a limited amount of time for activating these defining concepts and, furthermore, no conscious intent to activate them, then the positive defining concepts of special pairs will be more likely to be activated during both presentation and test than the negative defining concepts of ordinary pairs. Thus, there will be better replication of the memory trace and the retrieval process for special pairs than for ordinary pairs. One could also argue that the number of pairs in any special class is smaller than the number of pairs in the ordinary class, which makes the defining concept a more valuable cue for special pairs than for ordinary pairs.

This work was supported primarily by grant MH 08890-01, from the National Institute of Mental Health, U.S. Public Health Service. Further aid was received from grant MH 05673-03, to Hans-Lukas Teuber.

REFERENCES

KLEINKNECHT, H. (1906). The interference of optical stimuli. *Harvard psychol. Studies* **2**, 299–308.

OBONAI, T. & TATSUNO, C. (1954). Facilitation and inhibition in memorizing a series of digits or letters. *Jap. psychol. Res.* **1**, 1–8.

RANSCHBURG, P. (1902). Über Hemmung gleichzeitiger Reizwirkungen. *Z. Psychol.* **30**, 39–86.

TATUNO, T. (1961). On the effect of the isolated term and the identical term in serial learning. *Jap. psychol. Res.* **3**, 17–27.

TURLEY, L. A. (1906). Inhibition and reinforcement. *Harvard psychol. Studies* **2**, 293–7.

UNDERWOOD, B. J. & SCHULZ, R. W. (1960). *Meaningfulness and Verbal Learning*. Chicago: Lippincott.

WICKELGREN, W. A. (1965). Short-term memory for repeated and non-repeated items. *Quart. J. exp. Psychol.* **17**, 14–25.

WILLIAMS, J. P. (1961). A selection artifact in Rock's study of the role of repetition. *J. exp. Psychol.* **62**, 627–8.

(*Manuscript received* 14 *August* 1964, *and publication delayed owing to editorial oversight*)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 275–289
Printed in Great Britain

THE DEVELOPMENT OF A KEYBOARD TRAINING SITUATION IN MINIATURE

BY J. A. LEONARD*

Medical Research Council Applied Psychology Research Unit, Cambridge

In this paper, the author describes progressive changes which have taken place in a methodology for the study of keyboard tasks. The S–R relationships progressed from a laboratory abstraction to one having high face validity; practice was extended from a single session to programmes in which two sessions were given on each of ten days; and the relationship between the experimenter and the subjects became increasingly similar to that found in actual training situations. These changes are illustrated by examples drawn from the author's work. It is claimed that increasing the general face-validity of the laboratory approach may enhance the possibilities of studying 'pure' problems.

I. INTRODUCTION

Most research workers, at various stages of their work, have come up against the problems concerning the extent to which the tasks which they have chosen as experimental media influence not merely the specific results which they obtain, but much of the thinking about their subject-matter. There is, of course, the other equally interesting question as to why they have chosen a particular problem, and tackled it in the form of certain tasks—rather than other problems and other tasks—but that is another story.

It would, presumably, be generally agreed that in the laboratory, one is dealing with situations which are abstractions of real-life situations. That is to say, having decided to investigate a certain aspect of human performance, or of human conduct, one seeks to distil its essence, and make this concrete in a simplified, controllable form, suitable for experimental manipulation. It is necessary, therefore, to mention only in passing two points. It seems likely that one of the features which our present phase of development as a science will be remembered by, is the growing awareness of the extent to which the choice of experimental situations determines the generalizations which are subsequently put forward, not merely to the profession, but also to society in general. The second point is that there is always the possibility that experimental situations become 'functionally autonomous', in the sense that they have a built-in tendency to shoot off from their original setting, and create their own orbit and solar systems, with satellites performing an obedient rite, the whole procedure increasingly removed from the original intention.

It is intended to describe the manner in which, over the past ten years or so, the medium with which the author has been working, has undergone a gradual change. Some of the problems have remained, and a few others have been added. It is intended to demonstrate this change by reference to two fairly easily observable features, and one which is rather less so: first, there has been a change in the nature of the Stimulus–Response feature; secondly, there has been a change in the concept of 'practice';

---

* Now at Psychology Department, University of Nottingham.

and thirdly, there is the rather more elusive feature concerning the changed relationship between the subject and the experimenter.

It will be realized that the changes to be described are also reflected in the work of others, and that this is yet another manifestation of the *Zeitgeist*. But it is intended to use some of the author's old and more recent results, more by way of illustrating the changes, than in order to discuss specific issues.

II. CHANGES IN THE S–R RELATION

The author had become specifically interested in the problem of 'effector' anticipation, and happened upon an apparatus solution which enabled one to demonstrate one of the characteristics of all useful anticipation (Leonard, 1953); anticipation may allow one to perform more speedily, it certainly allows one to perform more smoothly, with one response merging into the next without any hesitation. The particular task was never used again subsequently, but the apparatus has been found convenient for other purposes, and its display-control set-up is shown in Fig. 1. The display of five lights was arranged in a pentagon (the reader should disregard the aperture in the centre at this stage) and the control consisted of five disks set in a corresponding pentagon, with a homing disk in the centre. The S–R relation was quite simple: to make a correct response, the subject had to move the stylus, which can be seen lying on the panel, from the homing disk to the corresponding disk: if the light at 12 o'clock was on, the move had to be to the disk at 12 o'clock, and so on: today one speaks of one-to-one matching, and high compatibility. In this type of apparatus the display would change whenever the subject made a correct response, and the basic unit of performance was a series of fifty stimuli. The scoring was in terms of times per unit, number of errors, and of paper records giving rather more detailed results about the constituents of each response.

An even simpler display of five lights in a row was used next, with five morse-keys to respond with. The display was now made to change whether the response was right or wrong, and the input included one in which there was a frequency bias, that is to say, one of the lights could be made to appear much more frequently than the others (cf. Leonard, 1958).

The next step in S–R complication was to make use of a simple, number inline display as shown in Fig. 2 on the left; now, the digit *one* required the subject to press the key on the left, the digit *two* the key next to it, and so on, the labels on the keys being visible to the subject. This is a slightly more difficult task, as will be seen later, but still one in which one display item only appeared at any one time.

Logically, though of course not chronologically, the next step in S–R compatibility was to choose a display which consisted of five pure tones, the subject having to respond to the lowest tone by pressing the key on the left (key one in Fig 2), the next highest with the next key to the right, and so on.

It is worth noting that this was found to be a difficult task to teach when an equiprobably input was used; it can be recommended to anybody who wants to study performance when, in the presence of complete objective certainty of the S–R pairing, subjects are liable to remain in a measurable state of subjective uncertainty for some considerable time.

Up to this there had always been a 'simple' one-to-one relation between display

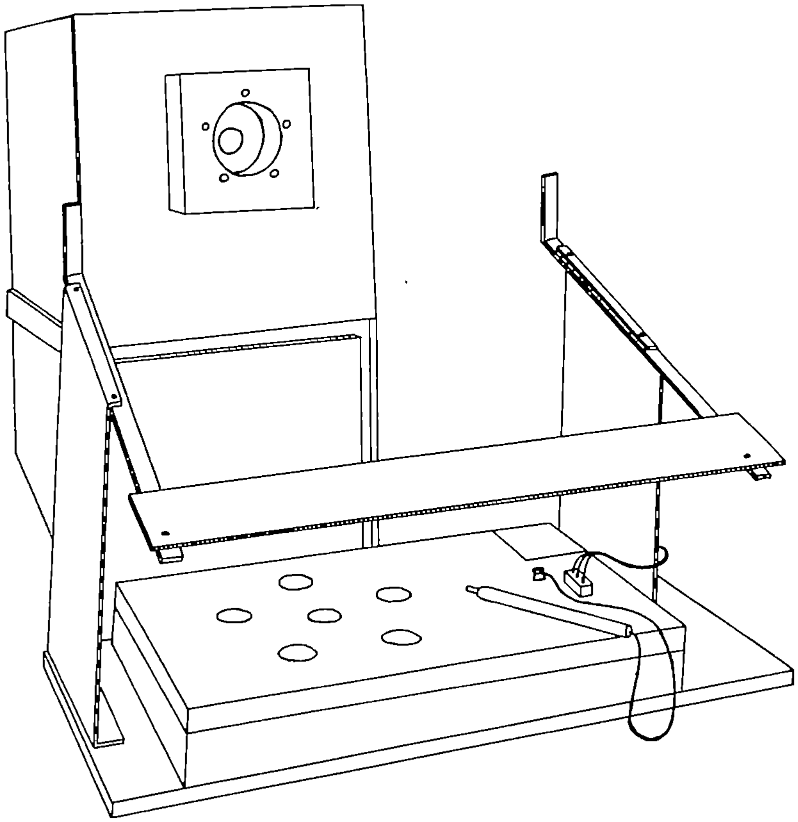


Fig. 1. View of five-choice apparatus. The positional display consisted of five lights surrounding the central aperture while the colour display was provided by rear-illumination of the bright patch shown in the centre of the aperture and slightly recessed.

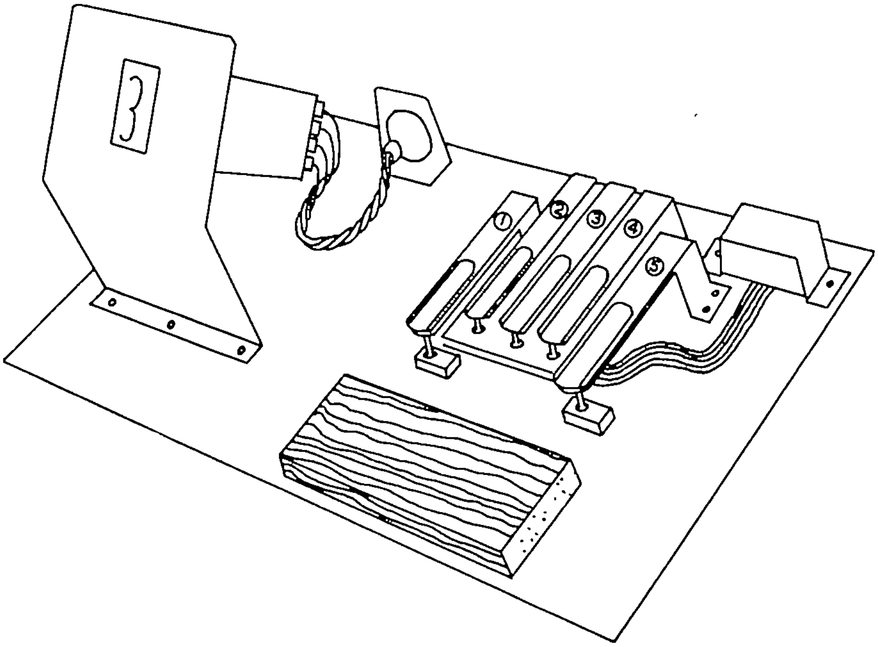


Fig. 2. Inline display and keyboard.

and control. There was only one departure from this in a study with the stylus apparatus, when a colour inline display was used (sited at the back of the centre aperture shown in Fig. 1) where the relation between colour displayed and 'correct' disk was rather arbitrary (cf. Leonard, 1964).

Again, up to now, the subject had always been presented with one display item at a time, no preview at all. Branching off, therefore, from the five-alternative number, inline task, printed strips were now used like that shown in Table 1. The S–R coding was still the same, and it will be noticed that here a heavily biased input was employed: there were on average about eight times more 'threes' than any of the other numbers. The main difficulty in this type of 'copying' task is for the subject not to lose his place in the list. This was overcome by drawing lines between groups of five digits, and providing the subject with an auditory *thump* after each fifth response; as well as with a re-setable counter which told him how many responses he had made.

Table 1. *Example of material used for the copying experiments*

(The input is heavily biased on source 3, there are dividing lines between successive groups of five items, and the end of each group is labelled to allow the subject to make use of the response counter.)

2	3	3	4	2	3	5	1	3	5	3	5	3	3	3	3	3	3	2	2	3	3	3	3	4
				5					*10*					*15*					*20*					*25*
4	3	5	3	3	3	3	5	5	3	3	1	4	1	3	3	4	3	2	4	3	3	3	3	3
				30					*35*					*40*					*45*					*50*

Table 2. *Example of list used for the word-copying experiment*

HANDS	ARSON
HOODS	OTHER
TRASH	SHEAR
ROAST	SNEER
SHOAL	HEATS

Now with this task one is approaching something like face-validity, in the sense that, on a very small scale, this might be said to resemble something like a real copying task: resembling just a bit more closely something that might take place outside the ivory tower. The next step from here was to go to meaningful words. Very fortunately, Dr R. Conrad had already done most of the spade work by the brilliant utilization of 'ADEHLNORST'. These are, with one exception, the ten most frequent letters of the alphabet in English usage and it is possible to make up vocabularies of all sorts of meaningful words from them. The use to which they were put, is shown in Table 2: using two keyboards, the keys were labelled ADEHLNORST, from left to right, and the input was 'grouped' in meaningful words of five letters each. To enable the subject to keep his place in the list there was a lamp opposite each word; the lamp would stay on until the last letter of a word had been keyed—and then the next lamp would come on: this method was used in preference to the 'thump' described earlier on. That is the story as far as the development of the S–R relation goes.

Table 3. *The experimental plan used for the most recent series of experiments*

Day 1

Session 1, a.m. — Session 2, p.m.

Block (1½ min. rest) Block

50 RESPONSES	50 RESPONSES	50 RESPONSES	50 RESPONSES	50 RESPONSES

Extent of practice

Days 1–10: Monday–Friday of 2 successive weeks

Each Trial:	50	Responses
Each Block:	250	Responses
Each Session:	500	Responses
Each Day:	1,000	Responses
Total:	10,000	Responses

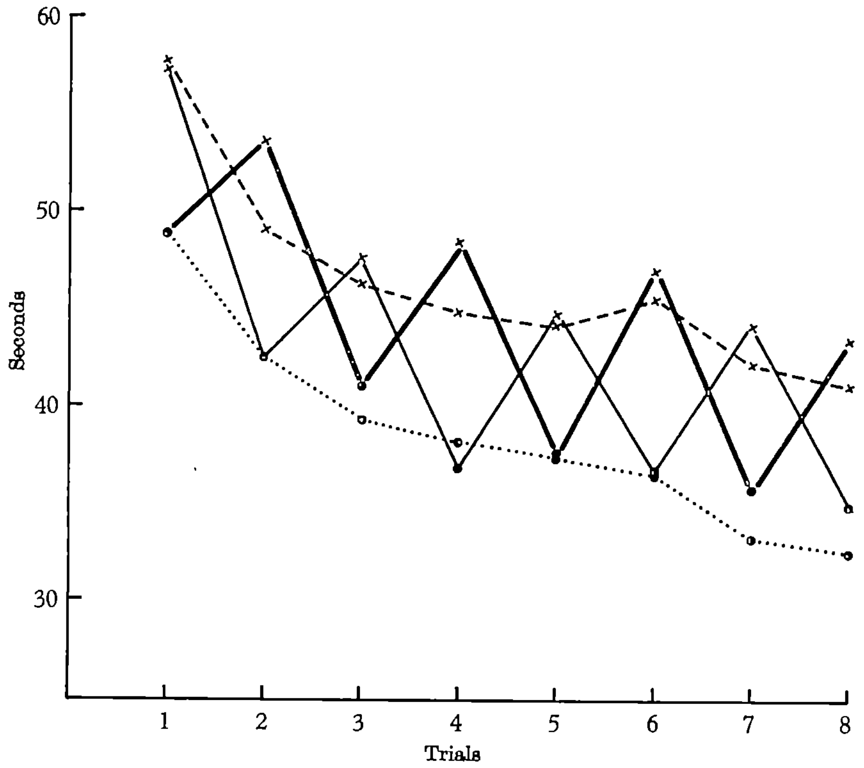


Fig. 3. Results of study of transfer (from Leonard, 1952). Mean times per trial ($n = 10$/group); ×—×, condition (1); ●—●, condition (2); ---, group A (control); ..., group B (control); —, group C (experimental); —, group D (experimental).

III. Changes in the concept of practice

On the practice side one could be much briefer, and confine oneself to the simple statement that one started off with the idea that asking subjects to make 200 responses during a single session, and comparing the first with the last fifty, would be quite adequate; more recently the standard design has been the arrangement shown in Table 3: subjects came twice a day for a fortnight, 10 working days. Each day they attended for two sessions. During each session they had to make a total of 500 responses, broken up into trials of fifty responses. One therefore had some 10,000 responses per subject, and much of the data analysis concerned itself with comparing effects over the last 3500 of these.

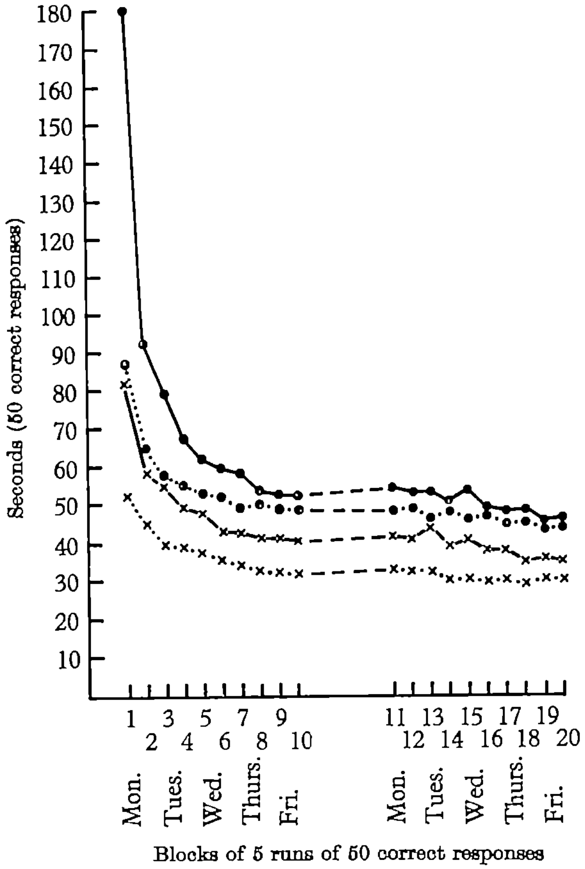


Fig. 4. Results from a five-choice coding study (from Leonard, 1964). Learning curves for four conditions of a coding study; each point is the average of 25 values (5 subjects × 5 runs): ●—●, colour per touch; ●---●, colour per vision; ×—×, position per touch; ×---×, position per vision. (Five subjects in each group.)

But this is perhaps the point at which to introduce some of the results obtained and to underline the change of medium in this way. Mostly, these results will be in terms of time per response or per trial, for generally errors were kept down to about 2 %. Groups of six subjects or so were used typically in designs in which each subject was exposed to one condition only. It should be pointed out, however, that for most of the critical conditions there are several repeat studies. Fig. 3 shows the results from an unpublished study (Leonard, 1952). The apparatus used was the old five-choice, and it was desired to show the differential effects obtained when transferring from a non-anticipatory to an anticipatory condition, and vice versa. There were two control groups of ten subjects each: the curve on top shows the performance for non-anticipation, and the one along the bottom, that for anticipation. And there were two experimental groups of ten subjects each: one starting with non-anticipation and, alternating with anticipation, and the other the other way round. To begin with, a clear differential effect was obtained, in the sense that there was complete transfer when going from non-anticipation to anticipation, but only partial positive transfer when going from anticipation to non-anticipation. But towards the end of practice this differential effect had vanished.

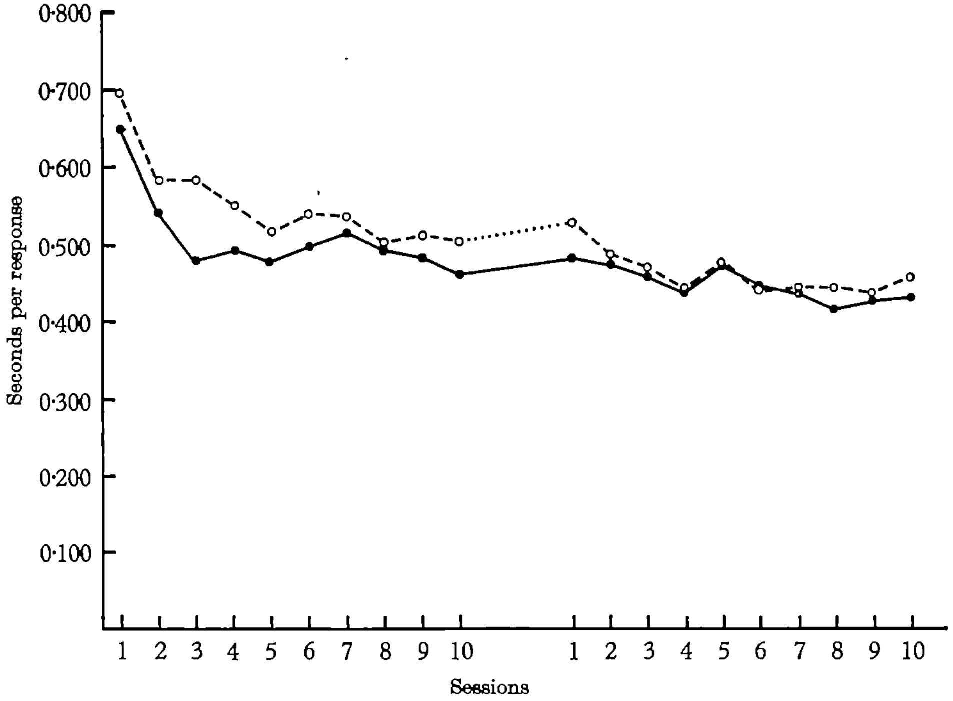


Fig. 5. Results from a study comparing a group of subjects performing once a day for 10 days with that of a group performing twice a day for 5 days: ● --- ●, 1 session per day for 2 weeks; ●—●, 2 sessions per day for 1 week.

'The end of practice', that is the point: there were, as can be seen in Fig. 3, eight trials of fifty responses each, given during just one experimental session. Fig. 4 shows the results from Leonard (1964). A modified version of the stylus apparatus was used to study the interrelation between S–R coding and the ability to observe movements of one's hand. For our present purposes it suffices to say that a rather large effect of precluding vision of the hand was observed early on in training, but that this effect had become much reduced by the end of practice. The main point is that Fig. 4 is a practice curve in which there are 10 days along the horizontal scale, with subjects producing at least 500 responses during one daily session.

This was still not the final form of the practice design: Fig. 5 shows results obtained during a lengthy series of pilot experiments on practice designs: the dotted line represents performance of a group of subjects who attended for one session daily, in two blocks of five trials of fifty responses each; the solid line shows the performance of another group the treatment of which was comparable to the first group's in all respects, but one: the subjects attended for two sessions daily, so that one now had 1000 responses a day. This figure is plotted block by block, and not on a common time scale as far as practice is concerned. The task, by the way, was a five-alternatives task with a 68 % frequency bias on one source; the display five neon lamps, and the control a keyboard. Now these results were interpreted in the sense that there was little difference, for this sort of task, between giving ten sessions spread over 10 or over 5 days. This meant that one could either confine testing to one week only, or get twice the amount of data, and use the second and presumably more 'stable' week to introduce changes in the experimental conditions.

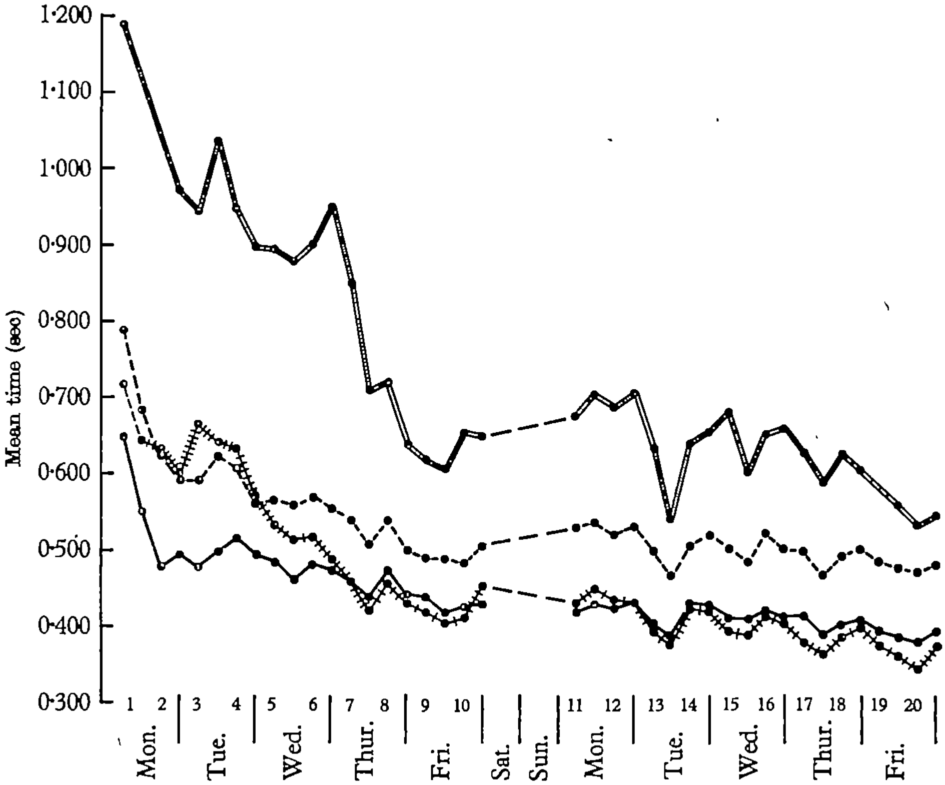


Fig. 6. Results from a study comparing four different types of display (from Leonard & Newman, 1965): ●=●, auditory; ●---●, numerical projected; ●—●, positional; ●╪●, numerical printed.

Fig. 6 shows how this design was used to compare some of the different displays mentioned earlier: there is the auditory display on top, coming to within hailing distance of the others towards the beginning of the second week. One further down can be seen the performance of the 'number inline' group, with the performance of the group responding to one of five lamps making up the simpler of the two bottom lines. The remaining curve is that of the 'copying' group, which starts off rather more slowly and then drops down to become comparable to that of the 'five lamps' group. It will be recalled that the copying group had the same S–R coding as the number inline group, but that there was full preview provided.

It should be noted that from about the second half of the first week onwards, the ups and downs on all the curves in Fig. 9 are in closely similar places for all groups—a point which is considered again later. Fig. 7 shows one way of varying experimental conditions during the second week. One group of subjects was given what have since become standard instructions: 'Perform as fast as you can afford to, without making more than one error per fifty responses on average'. In the case of the other group (the dotted lines in the figure) their instructions encouraged speed initially, without discouraging errors. Then, on the afternoon of the Tuesday of the second week, the second group was asked to accept the accuracy level of the first group; as can be seen, from the lower set of curves, the errors came down straight away (and did so for all subjects); over the remaining 3½ days performance was only slightly and not significantly slower than that of the group which had started with the stringent accuracy instructions. Replication with a different display condition gave a

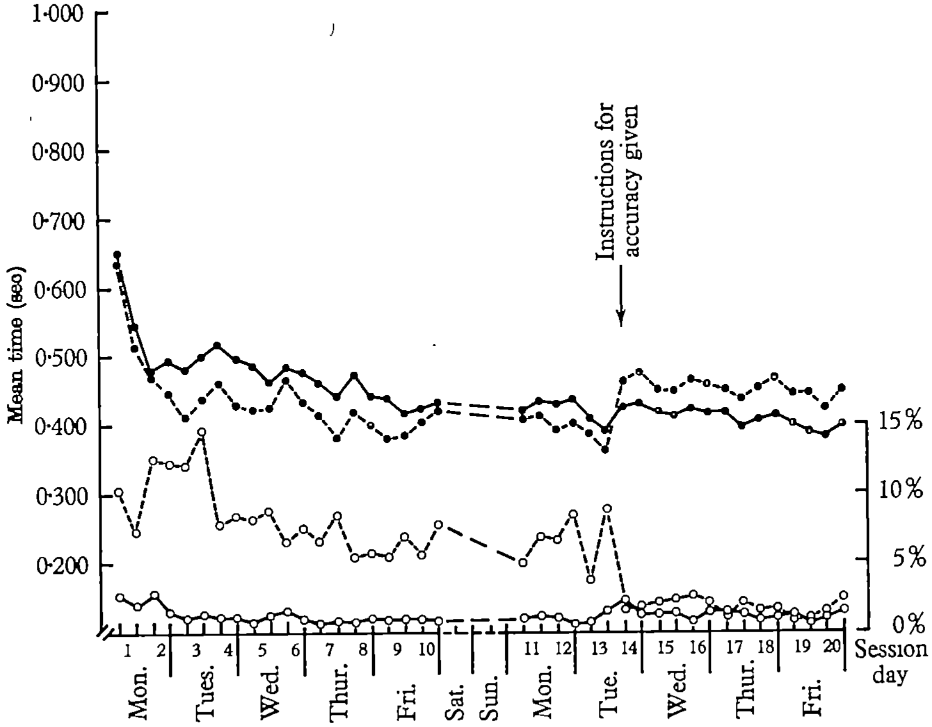


Fig. 7. Results from a study in which one group was instructed to maintain high accuracy throughout training, while another was allowed to make errors during the first part of the experiment (from Leonard & Newman, 1965.) Accuracy instructions: ●—●, mean times, ○—○, errors. Speed instructions up to 13th session: ●---●, mean times, ○---○ errors.

very similar result. (The results shown in Figs. 6 and 7 were reported by Leonard & Newman (1965).)

Fig. 8 shows the result of a study (Leonard & Newman, 1964*b*) on an old problem: how one can tell whether subjects are typing in terms of letters rather than in terms of some higher order 'habits' or units such as words. Taking the basic 'ADEHLNORST' vocabulary as the starting-point, it was argued that if one trained subjects on one sample of words and then transferred them to another, comparable set of words, one would expect some decrement in performance if the subjects had already learned to deal with the material in terms of words, that is to say, if they had learned to generate rather more complex and specific motor patterns; but as long as they were still dealing with the input letter-by-letter, there should be little or no effect. One group was trained on a 100 word vocabulary, and another on a 10-word vocabulary. Then, at the same point during the second week, each group was transferred to another sample of the corresponding size. As can be seen, there was no effect with the 100-word group (the thin, solid line) but there was an effect with the 10-word group.

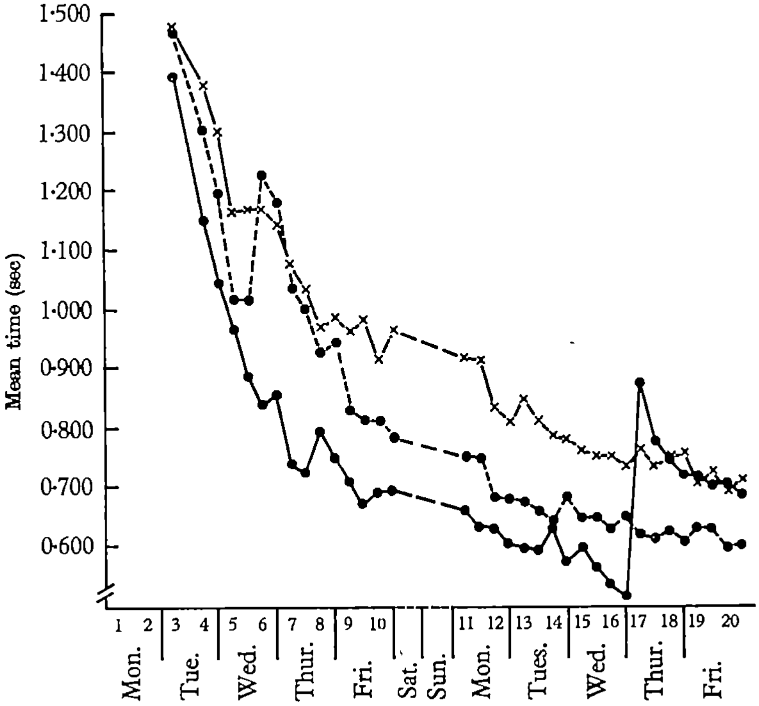


Fig. 8. Results from an experiment to study the acquisition of 'higher habits' (from Leonard & Newman, 1964*b*): ●—●, small vocabulary, late transfer; ● - - - ●, small vocabulary, early transfer; ×—×, large vocabulary, late transfer.

To conclude this section, a change in the manner of examining results can be illustrated. Fig. 9 shows the type of analysis which was much in vogue up to the late fifties: response distributions from which one can see the effect of a 'machine lag' of a certain duration when subjects had to deal with *either* an equiprobable (upper half of figure) *or* a heavily biased input (lower half of figure). As a comparison of the solid with the dotted lines shows, it turned out that with this particular lag only the subject having to deal with a heavily biased input suffered some decrement, though not as much as might have been expected (Leonard, 1958).

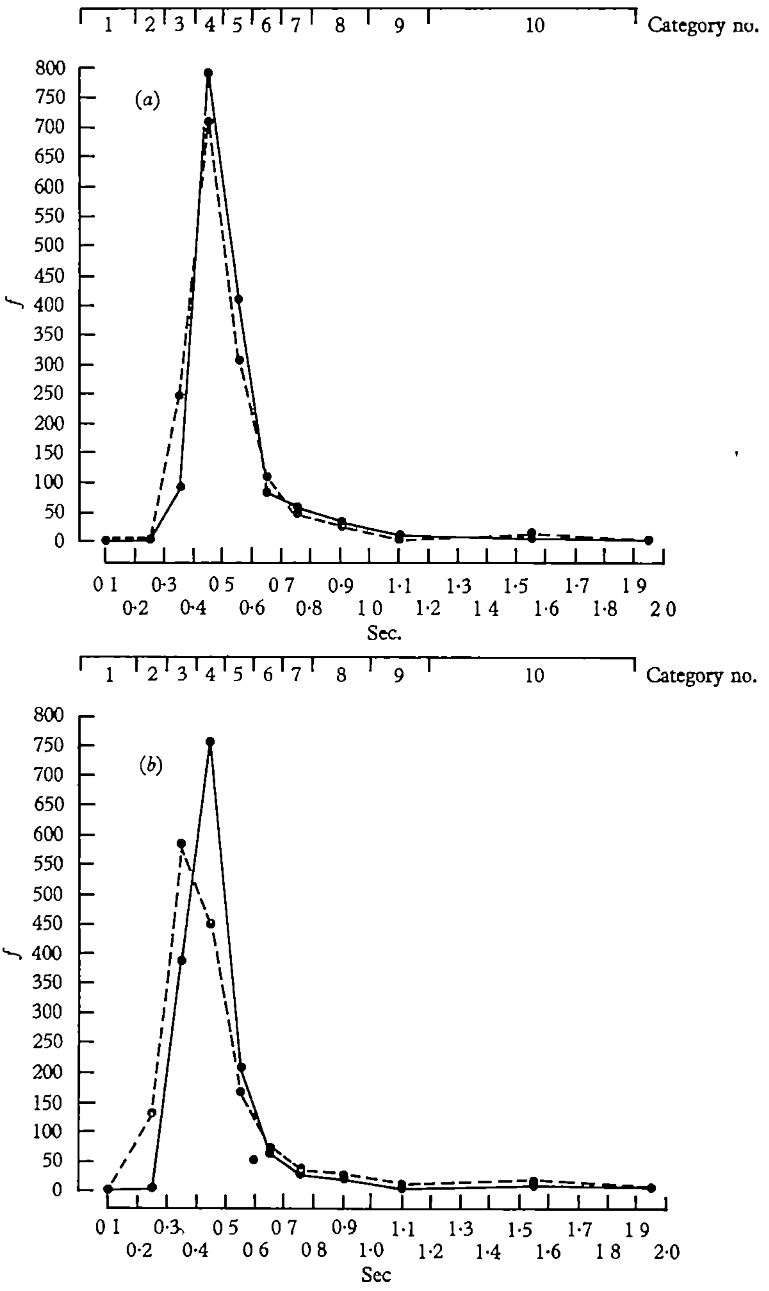


Fig. 9. Results of a study in which the effect of a machine lag was studied when subjects were responding to either an equiprobable or a frequency imbalanced input (from Leonard, 1958) Frequencies of responses (*f*) in different time-categories. (*a*) Balanced series, (*b*) biased series. Dotted lines: without enforced delays in responding. Solid lines: with enforced delays.

Fig. 10 has a more contemporary look. In Fig. 9, in which four different display conditions were compared, there were ups and downs which were in closely similar places for all four curves. There is evidence now that these variations are mainly related to fluctuations of the frequency bias in the input. A basic sequence of 10,000 stimuli having an average bias of 68 % on source three was used. When this sequence is broken up into trials of fifty, there will be some trials with an above-average and some trials with a below-average bias. In a recent study the details of how subjects 'follow the bias' was examined by carrying out analyses of successive responses.

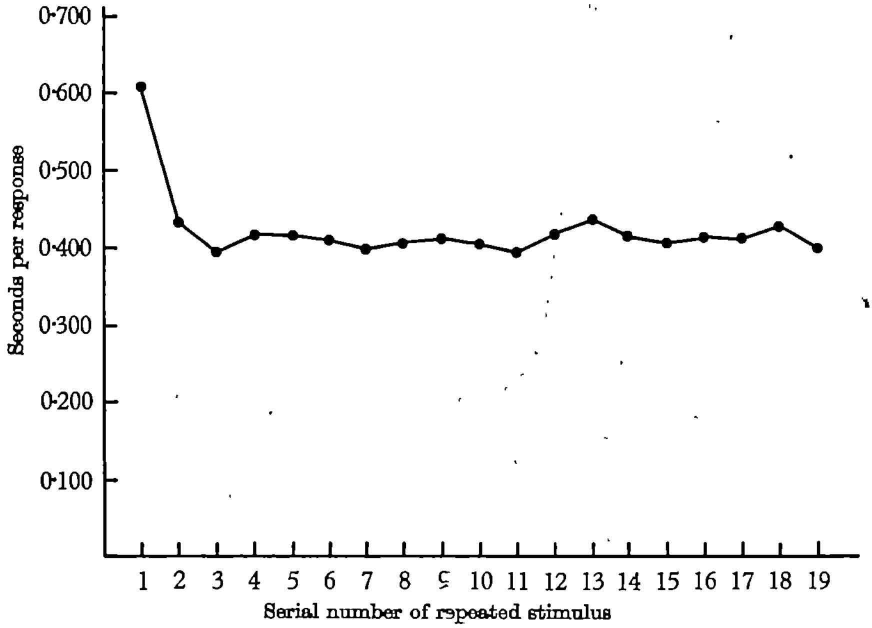


Fig. 10. Changes in response time when subjects had to respond to a very long run of the same stimulus embedded in other stimuli.

Fig. 10 shows what happened when subjects had to respond to a 'run' of nineteen consecutive 'threes' during one of the trials: they were rather slow on the first response but settled down pretty smartly to a fairly steady level (cf. Leonard, Newman & Carpenter, 1966).

IV. THE RELATION BETWEEN EXPERIMENTER AND SUBJECTS

To wind up this section, a few words about what appeared to be necessary in the relationship between experimenter and subject seem indicated. In the experimental set-up described, the subject sits inside a 'noise attenuating' cubicle and the experimenter outside with the noisy equipment. Subjects are watched through a one-way vision screen. The subject knows that he is being watched, but having the screen allows one to move about without distracting the subject. The experimenter talks to the subject through a one-way intercom; the subject still talks back by the old method of raising his voice. Most of the conversation is one-way: at the end of each trial the subject is told the time per trial and the number of errors made. In most experiments this has been all the knowledge of results given to the subject.

One tries to persuade the subjects to maintain a constant aim: to attempt to increase their speed without lowering their accuracy. This means, among other things, that some over eager subjects may have to be quietened down, while the more timid may have to be encouraged. What has been most surprising is that subjects, mostly sailors, did not seem to find this 10-day schedule as boring as one might think. On the whole, co-operation was good, and was shown by the fact that good performances were still obtained on the last Friday afternoon which was followed by week-end leave. What is more, there were some subjects, usually the better ones, who really enjoyed what they came to regard as a challenge; and at the other end of the scale there were those who arrived firmly convinced that they were no good at this sort of thing, or indeed at anything, and who were sometimes touchingly grateful by the end of the fortnight because they had been brought to levels of performance which were respectably close to the group average. Between-subject rivalry was hardly ever encouraged, but within-subject improvement was stressed. In short, it is *not* thought to be useful in this type of experiment to give subjects a 'standard' treatment throughout, and that it is worth taking the risks inherent in this attitude. The only ground on which subjects were rejected was detected apparatus failure, a drop-out of more than one day during practice, or previous experience with the experimental or a similar task.

V. SOME SPECIFIC CONCLUSIONS

This account covers roughly a decade of work. Some of the more specific conclusions are summarized before returning to the main discussion. As far as the study of different S–R relations is concerned, it has been possible to show differences between conditions which appeared to resist the effects of practice as far as it was practicable to take it: even the apparently simple numerical inline display could not be brought to the level of the five light, positional display. It is true that the same subjects have never been used on both codes, but sufficient data from many groups of subjects has been obtained to make the experimenter feel confident that much more practice than it was possible to give would be needed to make the difference between these two codes disappear. The persistent difference is more easily acceptable in the case of the colour and the auditory codes.

Without explicitly intending to do so, the work has become increasingly involved in the problem of knowledge of results and the conditions under which subjects can rely increasingly on an internalized knowledge of results. Although it has not been possible to put this to a direct formal test as yet, sufficient evidence has been accumulated enabling one to specify a point during learning at which external knowledge of results can be withdrawn. It is suggested that this point is reached when the subject, given as much time as he likes, can produce the adequate response to a given stimulus with a high degree of reliability.

Lastly, it is submitted that the foundations have been laid for a more comprehensive study of various aspects of the 'coding uncertainty' (Leonard, 1961), which arises from the fact that almost all the S–R relations which are used, both in the laboratory and in the outside world, are arbitrary and have to be learned. It is possible to break down this coding uncertainty into three components: uncertainty about the stimulus array, about the response array, and about the relation which exists between stimuli and responses. And it is now more possible to manipulate these independently: the auditory display, for instance, provides maximum uncertainty about the nature of the display; preventing the subject from observing his movements, and using graded rather than discrete responses introduces greater uncertainty about the response; and randomization of the relation between stimuli and responses, particularly when the randomization is not a stable one, increases this aspect of coding uncertainty.

VI. DISCUSSION

It is submitted that over the years an experimental situation has been developed which has become increasingly life-like. One of the built-in advantages of this is that, allowing for the scaling down of inputs and of training, one can say something about the level of performance of subjects, with respect to that obtained by people who have to make their living by pounding keyboards; in the laboratory one finishes up very approximately where economically viable existence begins. That, to the author at any rate, is a sobering thought.

During the course of this work there have been two incursions into what might be called the applied field; both of them might eventually lead to implementation. A significant correlation between one version of the inline number task and proficiency in typewriting training was established (Leonard & Carpenter, 1964) and another version was used to illustrate some possibilities of automating keyboard training (Leonard & Newman, 1964*a*). It is certain that by carrying out this programme a better understanding has been gained of some of the problems which face the practitioners, administrators, instructors, and trainees, in the keyboard world. One aim of this paper has been to show that by increasing the face-validity of a task one need not necessarily preclude oneself from continuing to attempt to deal with 'fundamental' or 'pure' problems. Indeed, it is held that it is precisely by increasing face-validity, by coming closer to real-life situations, and by being able to relate the performance levels obtained in the laboratory to those obtained in real life, that one has *one* route towards a better understanding of human performance—and perhaps of human conduct.

This work has been going on for many years and is the result of collaboration of quite a number of people. Statistical help has been received from Dr Stone, Mr Freeman and Mr Ross; apparatus help from Dr Carpenter. For the past four years much of the really hard work has been done by Mr Newman. Thanks are also due to the Royal Navy, for allowing the use of subjects when available from other duties, and to the Medical Research Council. Much encouragement and great freedom has been received to pursue this approach, from the two successive directors of this unit, Dr Mackworth and Dr Broadbent. But thanks are also due to the founder of the unit under whom this work began: Sir Frederic Bartlett.

REFERENCES

LEONARD, J. A. (1952). Some experiments on the temporal relationship between information and action. Unpublished Ph.D. thesis, University Library, Cambridge.

LEONARD, J. A. (1953). Advance information in sensori-motor skills. *Quart. J. exp. Psychol.* **5**, 141–9.

LEONARD, J. A. (1958). The effect of 'machine lag' on a serial choice task with balanced and biased input frequencies. *Ergonomics* **2**, 44–51.

LEONARD, J. A. (1961). Choice reaction time experiments and Information Theory. In *Proc. 4th Symp. on Information Theory*. London: Butterworth.

LEONARD, J. A. (1964). On the relationship between coding and the ability to see what one is doing in a sensori-motor skill. *Quart. J. exp. Psychol.* **16**, 325–31.

LEONARD, J. A. & CARPENTER, A. (1964). On the correlation between a serial choice reaction task and subsequent achievement at typewriting. *Ergonomics* **7**, 197–204.

LEONARD, J. A. & NEWMAN, R. C. (1964*a*). 'Automating' training in a simple keyboard task. *Ergonomics* **7** (3), 305–309.

LEONARD, J. A. & NEWMAN, R. C. (1964*b*). Formation of higher habits. *Nature, Lond.* **203**, (4944), 550.

LEONARD, J. A. & NEWMAN, R. C. (1965). On the acquisition and maintenance of high speed and high accuracy in a keyboard training task. *Ergonomics* **8**, 281–304.

LEONARD, J. A., NEWMAN, R. C. & CARPENTER, A. (1966). On the handling of heavy bias in a self-paced task. *Quart. J. exp. Psychol.* **18**, 130–41.

(*Manuscript received* 31 *January* 1965; *revised manuscript received* 31 *March* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 291–295
Printed in Great Britain

THE EFFECT OF LEARNING INSTRUCTION ON MOTOR AND VERBAL RESPONSES IN RECALL

BY FITZ J. TAYLOR
Hatfield College of Technology, Hatfield, Herts

An experiment was designed to compare the effect of instruction (set) on verbal performance (recall) on the one hand and on overt behaviour (application) on the other. The experiment, using six groups of twenty subjects and three types of instruction procedures, showed that (*a*) a set-to-recall led to high recall-scores but low application scores; (*b*) a set-to-apply led to high application-scores and low recall-scores; (*c*) most surprising of all a learning instruction by which subjects were asked merely to learn without any indication of the purpose for learning resulted in significantly lower application and lower recall scores than either of the other two sets. Thus it seems that even a wrong instruction can sometimes be better than no instruction at all.

Eunice Belbin (1956) showed that information which has been learned can become apparent in at least two forms. It may emerge in ideational form such as recall or recognition or it may manifest itself in subsequent activity. The question as to what determined the preferred form for a given individual led Belbin to hypothesize that the manner in which information is retained (either in memory or for use in relevant action) is determined at the time of learning and conditioned by previously learned responses. Postman & Jenkins (1948) demonstrated that different learning instructions resulted in significantly different retention scores. It still remains, however, to put Belbin's view to the test and see whether a set towards learning in-order-to-use would facilitate overt behaviour while a set towards learning in-order-to-recall renders a subject better able to recall. Accordingly, an experiment was designed to look into this problem. Two hypotheses were set up: first that recall performance would be optimal if preceded by a recall set at the time of learning; secondly that overt-behavioural performance (application) would be optimal if preceded by a set-to-use the information.

METHOD

The plan of the experiment is shown in Table 1. The experiment was divided into two parts, each employing sixty subjects randomly allocated to three groups of twenty; subjects were tested individually. In each part group A were given 'application' instructions, group B were given 'recall' instructions and group C were given 'learning' instructions. In Part I all subjects were tested for application; in Part II all subjects were tested for recall.

The task involved instructions for operating a dotting machine, and actual operation of the machine. This piece of equipment comprised a wooden box with a slot towards the left-hand side of the upper surface; a roll of paper tape is rotated by a motor so that the tape slowly passes across the slot.

Subjects. The experiment was carried out on 120 adult subjects consisting of postgraduate and undergraduate students, technicians and secretaries from various departments at Birkbeck College.

Procedure: Part I

Group A. Subjects in this group were given the *application instructions:* 'I am going to give you a short passage to learn. This passage describes an apparatus used in psychology. You will have five minutes to learn it. After the five minutes are up I will give you a test on the apparatus to see how well you apply what you have learned. Is that clear? You may begin.' Each subject was then presented with a learning sheet which read:

'The dotting machine is an instrument used by psychologists. It has been employed in the study of skills, attention, personality types and perceptual functions. It is a small piece of equipment comprising principally, a wooden box (7·5 in. × 12·5 in. × 7·5 in.) with a slot (1·0 in. × 2·8 in.) towards the left hand side of the upper surface. A roll of paper tape is attached and is rotated by a motor so that the tape slowly passes across the slot.

Table 1. *The plan of the experiment*

Group	Instructions	Test
	PART I	
A	Application	Application
B	Recall	Application
C	Learning	Application
	PART II	
A	Application	Recall
B	Recall	Recall
C	Learning	Recall

'During the experiment the individual is asked to mark the circles which appear on the tape as it passes the slot. The more he is able to mark the higher will be his score. It has been found that the best way to mark the dotting machine is firstly to stand before it; secondly, to use a pencil in preference to any other writing implement; thirdly, to keep within the limit of the slot which is marked 0·5 to 1·5; fourthly, to keep your preferred hand in contact with the spot marked "hand"; and lastly, to keep the free hand in a fixed position, preferably behind your back.'

Immediately after the learning session each subject received the *application test instruction:* 'In this cubicle you will find a dotting machine. I would like you to operate it so as to get the highest score you can. You have two minutes. Is that clear? Now you may begin.'

While the subject performed on the machine his application score was recorded. The five points recommended in the final paragraph of the learning sheet were used as the criteria of application. The experimenter observed and noted whether the subject put these points into practice. The test lasted for two minutes and was divided into twelve 10 sec. intervals. The subject was awarded one mark for each of the five points that he applied continuously throughout a 10 sec interval. Hence, for the test as a whole, a subject could receive a maximum of twelve marks for each of the following: (*a*) standing while operating the machine (a chair was placed in position for the machine); (*b*) using a pencil (the subject had a choice of the pencil, a black ballpoint and a red ballpoint); (*c*) keeping his marks within the indicated limits 0·5–1·5; (*d*) keeping the operating hand in contact with a spot marked 'hand' on the machine; (*e*) keeping the free hand in a fixed position during the testing session. In addition the subject was awarded one mark if he recognized the dotting machine from two alternatives when he entered the cubicle. The total possible score was therefore 61 marks.

Group B. Subjects in this group were given the *recall instructions:* 'I am going to give you a short passage to learn. This passage describes an apparatus used in psychology. You will have five minutes to learn it. After the five minutes are up I will give you a short paper-and-pencil test to see how much of it you can remember. Is that clear? You may begin.' Each subject was then handed the same learning sheet as group A. After the learning session the procedure was identical to that for group A.

Group C. Subjects in this group were given the *learning instructions:* 'I am going to give you a short passage to learn. This passage describes an apparatus used in psychology. You will have five minutes to learn it. Is that clear? You may begin.' This was again followed by the same procedures as for the other two groups. All three groups therefore were tested on the dotting machine.

Procedure: Part II

Group A. This group received the ***application instruction*** given to group A in Part I of the experiment. They were then presented with the learning sheet. Next they were handed the following questionnaire and told: 'I would like you to complete this questionnaire. You have five minutes.'

NAME .... AGE............ SEX .... COURSE .... YEAR ......

1. What is the name of the instrument described in the passage?
2. Within what limits is it advisable to keep your markings on the machine?
3. Can you complete the following sentences?
   (*a*) It has been employed in the study of....
   (*b*) It is a small piece of apparatus comprising principally....
   (*c*) During the experiment the individual is asked to....
4. What is the size of the slot mentioned in the passage?
5. What is the exact location of the slot?
6. What is the size of the box?
7. Describe the best ways of performing on this instrument.

Group B. This group was given the ***recall instructions*** given to group B in Part I, followed by the learning session. They were then given the questionnaire to complete.

Group C. This group had the same ***learning instructions*** as group C in Part I, followed by the learning session and the recall questionnaire.

The questionnaire was scored in the following way. One mark was awarded for a correct answer to question 1. Two marks were possible for question 2, one for each dimension correctly given, i.e. 0·5 and 1·5. The three subsections (*a*), (*b*) and (*c*) of question 3 were devoted to verbatim recall, and one mark was given for each word correctly reported; the maximum possible marks were seven for (*a*), twenty for (*b*) and fourteen for (*c*). Two marks were possible for question 4, one for each dimension correctly given, i.e. 1 in. and 2·8 in. Two marks were possible for question 5, one for a reply describing which surface of the box contained the slot and one for indicating in which section of the surface it was located. Three marks were possible for question 6, one for each dimension correctly given, i.e. 7·5 in., 12·5 in., and 7·5 in. Question 7 was devoted to recall of 'substance' and one mark was awarded for a correct report of each of the five suggestions in the learning sheet. The maximum score possible was therefore 56 marks.

The form of the questions was such that numbers 2, 3(*c*), 5 and 7 tested knowledge necessary for the proper use of the machine. The other questions tested knowledge which, although interesting, did not necessarily help the subject to operate the machine. The difference in types of question made it possible to examine Postman's hypothesis that subjects attempt to single out and to emphasize those aspects of learning which will be of use in test situations.

RESULTS

The average application scores and recall scores are set out in Table 2. In the design there are three systematic sources of variation: the instructions, the tests and the interaction between tests and instructions. A comparison between the variation due to application test and the recall test would, by the nature of the design, be non-informative, since two separate and distinct test measures were used. In treating the results each set of scores has therefore been standardized. The analysis of variance of the standardized scores is given in Table 3.

Instructions *per se* do not make a significant difference. It is the interaction between learning instructions and testing procedure that provides the greatest influence on the performance score. This means that efficiency in performance, be it in terms of memory or in terms of overt behaviour, varies significantly with the relation between instructions and testing procedure. Comparing the results obtained with different sets for each of the two testing procedures, reveals more clearly the nature of the interaction process.

Part I: application tests. Under this testing procedure both the learning and the recall instructions led to poorer performance than the application instruction, the learning instruction producing the poorest performance (in both bases $P < 0{\cdot}001$; in calculating the individual t tests an improved estimate of variance based on the error variance of the total distribution was employed). However, there was no difference between subjects instructed to recall and those instructed to learn ($t = 0{\cdot}5$; D.F. 38; $P > 0{\cdot}20$).

Table 2. *Average application scores and average recall scores for each of the three instructions; the maximum possible application score was* 61 *and the maximum possible recall score was* 56; $n = 20$ *for each of the six groups of subjects*

	Instructions					
	Application		Recall		Learning	
Test procedure...	Mean	S.D.	Mean	S.D.	Mean	S.D.
Application	27·05	2·52	20·50	6·62	19·70	9·50
Recall	27·30	6·90	31·05	4·31	25·10	7·56

Table 3. *Analysis of variance of the standardized scores*

Source	S.S.	D.F.	M.S.	*F*	*P*
Instructions	8·48	2	4·24	0·95	—
(Test)	—	(1)	—	—	—
Interaction	8·92	2	4·46	5·57	< 0·005
Residuals	89·96	112	0·80	—	—
Total	107·36	117	—	—	—

Part II: recall tests. With this procedure subjects instructed to recall did significantly better than those instructed either to apply or to learn ($t = 2{\cdot}55$; D.F. 38; $P < 0{\cdot}01$; and $t = 4{\cdot}05$; D.F. 38; $P < 0{\cdot}001$ respectively); and the learning instruction resulted in the lowest level of performance. Again there was no difference between subjects instructed to apply and those instructed to learn ($t = 1{\cdot}50$; D.F. 38; $P > 0{\cdot}05$).

Discussion

The results suggest that similarity between learning instructions and tests tends to maximize efficiency. Yet contrary to expectation, a learning instruction by which subjects were asked to learn without being given a reason for learning produced the most inefficient performance. As this is, in practice, equivalent to the absence of set, it throws further light on the nature of the interaction procedure. Whereas in the first place a high degree of similarity between instruction and test facilitates efficiency and dissimilarity impairs it, it is now seen that a neutral relation, i.e. subject learning without any thought of test, leads to the greatest impairment of all. One could therefore postulate a curvilinear relation between similarity of instruction and test on the one hand and efficiency of performance on the other.

Postman & Senders (1946) argue convincingly that even without specific instructions subjects bring to the learning situation their own self-instructions. These are the results of 'early established general sets'. In view of this, and the fact that the groups were random, it would be expected that the group of subjects who were instructed to learn without an indication of the type of test to follow would include some whose 'early established general set' would be toward application and others whose early general set would tend toward remembering. The 'appliers' in the group should tend to make the mean application score higher than that of the group instructed to recall. Similarly, the memorizers in the learning group should push-up the mean recall scores above that of the application group. In both cases this was not borne out.

A hypothesis proposed by Postman & Jenkins (1948) suggests that as a result of instruction, subjects single out and emphasize those aspects of the learning task which will be used in the test situation. This led us to analyse the recall scores in further detail. Whereas the recall group did significantly better than the application group on questions 1, 3*a*, 3*b*, 4 and 6 (in each case $P < 0{\cdot}05$ for Mann–Whitney U tests), the application group proved equally good at answering questions 2, 3*c*, 5 and 7. Reference to the wording of these questions shows that the application group did better on items which would be useful to them in an application test. This is oblique support for Postman and Jenkins' idea.

I acknowledge the facilities that I had in the Department of Psychology at Birkbeck College, London where the experiment was carried out.

REFERENCES

BELBIN, E. (1956). The effects of propaganda on recall, recognition and behaviour. *Brit. J. Psychol.* **47**, 163–73.

BELBIN, E. (1956). The conditions which determine the response to propaganda. *Brit. J. Psychol.* **47**, 259–70.

POSTMAN, L. & JENKINS, W. (1948). An experimental analysis of set in rote learning: the interaction of learning instruction and retention performance. *J. exp. Psychol.* **38**, 683–9.

POSTMAN, L. & SENDERS, V. L. (1946). Incidental learning and generality of set. *J. exp. Psychol.* **36**, 153–65.

(*Manuscript received* 15 *August* 1965; *revised manuscript received* 15 *March* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 297–300
Printed in Great Britain

THE RESPONSES OF NORMAL ADULT SUBJECTS TO THE MATRICES TEST

BY N. E. WETHERICK*
Medical Research Council Unit for Research on Occupational Aspects of Ageing, University of Liverpool

Bromley (1953) drew attention to the fact that what he called 'primitive forms of response to the matrices test' could frequently be observed in a sample of thirty-five hospitalized psychiatric patients between 40 and 80 years of age. He concluded that the theory on which the matrices test is based does not adequately account for the performance of 'primitive' subjects. He supposed the existence of a kind of abstract intelligence which is measured by the matrices test and is dominant in normal 'non-primitive' subjects but which overlies other primitive thought processes which may be or become dominant in primitive subjects.

The subjects in Bromley's sample consistently chose responses other than the correct response, which made it look as if they were using a consistent but incorrect method of solution—a primitive thought process. The grounds on which such processes are judged 'primitive' are that they have been observed to occur in elderly, deteriorated subjects. Bromley seems to have assumed that primitive responses of this kind do not occur in normal subjects. Analysis of the results of a sample of 236 normal adult males provided an opportunity to test this assumption.

The sample consisted of between forty-five and fifty subjects in each decade age-group 20–29, 30–39, . . ., 60–69 and was stratified by social and occupational class to form a representative cross-section of the male population of Merseyside. The usual negative relation between age and matrices score was present ($r = -0{\cdot}485$), but the average score obtained was higher than would be expected in a random sample, probably because the subjects were all volunteers. Ninety-nine of the 236 subjects were at or above the ninetieth percentile for their age-group and only twenty-four were below the fiftieth percentile. The most direct comparison possible is that between Bromley's sample and the forty-five subjects aged 60–69 yr in our sample. Bromley's sample had a mean score of 21·3 (S.D. = 9·2), ours a mean score of 34·8 (S.D. = 10·4). Our sample was of considerably above average intelligence; Bromley's was of below average intelligence, suffering in addition from some form of psychiatric illness requiring in-patient treatment.

A comparison of the proportion of occasions on which each possible response was chosen to each matrix reveals very close agreement between our sample of $n = 236$ and Bromley's sample of $n = 35$. Considering first the rank order of difficulty of the matrices in each set, there is a slightly closer relation between the orders for the two samples than between either sample and the putative order. If we consider the order of preference among possible responses to each matrix this is the same in almost

---

* Now Lecturer in Psychology, Department of Social Sciences, University of Bradford.

Table 1. *Frequencies of responses to the items of Raven's Progressive Matrices* (1938) *for a sample of* 236 *normal adult males*

(Bold-faced numbers are frequencies for correct answers; italic numbers are frequencies > 10% of the sample for other responses. The results for Bromley's (1953) sample of 35 'primitive' thinkers are given in brackets for comparison.)

Set A...	1	2	3	4	5	6	7	8	9	10	11	12
Alternative 1	—	—	**229** (32)	—	—	—	1 (1)	6	**231** (26)	—	*44* (3)	6 (3)
2	—	—	2	**229** (32)	1	—	2	**220** (27)	—	—	1	2 (4)
3	—	—	4 (3)	5 (1)	—	**234** (33)	1	— (2)	1	**228** (20)	3 (1)	3 (2)
4	**236** (35)	—	—	1 (2)	2	1 (1)	3 (6)	3 (2)	1	1 (2)	10 (11)	**220** (17)
5	—	**236** (35)	1	1	—	—	— (1)	—	3 (5)	1	**160** (6)	5 (1)
6	—	—	—	—	**233** (35)	1 (1)	**229** (27)	7 (3)	— (3)	6 (12)	18 (13)	— (7)
Not attempted	—	—	—	—	—	—	—	— (1)	— (1)	— (1)	— (1)	— (1)

Set B...	1	2	3	4	5	6	7	8	9	10	11	12
Alternative 1	—	1	**228** (26)	—	**229** (20)	— (4)	10 (7)	— (1)	*48* (17)	22 (9)	5 (3)	4 (1)
2	**235** (33)	2 (3)	6 (6)	**222** (24)	— (2)	13 (7)	1 (1)	6 (4)	2	19 (6)	23 (9)	*33* (0)
3	—	3 (1)	— (1)	2	—	**216** (16)	1 (1)	4 (4)	1 (1)	**192** (5)	*30* (4)	23 (3)
4	1	—	—	11 (5)	1	1 (2)	2 (1)	1 (1)	**157** (3)	—	**154** (3)	*25* (2)
5	—	—	1	— (2)	6 (4)	2 (1)	**194** (17)	3 (1)	24 (7)	2 (3)	7	**120** (1)
6	—	**230** (29)	1	1 (2)	—	4 (2)	*28* (3)	**222** (19)	4	1 (2)	17 (1)	*29* (5)
Not attempted	— (2)	— (2)	— (2)	— (2)	— (3)	— (3)	— (5)	— (5)	— (7)	— (10)	— (15)	2 (17)

Set C...	1	2	3	4	5	6	7	8	9	10	11	12
Alternative 1	—	— (2)	1 (3)	— (3)	—	9 (5)	6 (2)	**160** (5)	*28* (3)	13 (2)	**124** (1)	6 (1)
2	2 (1)	**228** (22)	3 (4)	1 (3)	1	3 (3)	—	8 (2)	2 (1)	8 (2)	10 (2)	**44**
3	1 (1)	3 (7)	**222** (15)	8 (4)	2 (5)	1 (3)	—	—	1	—	6 (1)	*91* (3)
4	—	2	1 (1)	20 (6)	10 (6)	**200** (4)	7 (3)	21 (1)	2	19 (2)	8 (2)	*64*
5	—	—	3 (1)	2 (1)	3	6 (3)	**212** (8)	3	4	5	*62* (1)	8
6	1	— (2)	3 (4)	2 (2)	—	—	3 (1)	10	21 (2)	**139** (1)	1	2
7	1	—	3 (4)	1	**216** (12)	2	3	20 (2)	**171** (4)	1	13	5 (1)
8	**231** (31)	3	—	**202** (11)	4 (4)	14 (2)	5 (1)	10 (1)	5	*47* (1)	7	9 (1)
Not attempted	— (2)	— (2)	— (3)	— (5)	— (8)	1 (15)	— (20)	4 (24)	2 (25)	4 (27)	5 (28)	7 (29)

Table 1 (*cont.*)

Set D...	1	2	3	4	5	6	7	8	9	10	11	12
Alternative 1	2 (3)	—	4 (3)	1 (2)	4	6 (2)	4	8	**180** (4)	3	4	5
2	—	6 (5)	2 (3)	4 (1)	4	3	3	7	11	**136** (1)	23	*29*
3	**222** (27)	5 (8)	**209** (11)	—	3 (1)	8 (2)	4 (1)	7	7	3	14	11
4	3 (1)	**214** (12)	10 (8)	3 (2)	4	2	20 (1)	**192** (5)	8	7	2	15 (1)
	1 (1)	1	—	14 (2)	1	4 (1)	**177** (6)	8 (3)	6 (1)	*43* (3)	**91** (2)	6
6	1	1 (2)	3 (1)	3 (1)	1	**203** (6)	—	1	3	5	*25*	**47**
7	1	2 (3)	— (1)	**202** (8)	1	4 (1)	6 (1)	5	3	5	*37*	*43* (2)
8	1 (1)	2	4	5 (1)	**213** (11)	1	16	2 (1)	9	23 (1)	*24*	*63*
Not attempted	5 (2)	5 (5)	4 (8)	4 (18)	5 (23)	5 (23)	6 (26)	6 (26)	9 (30)	11 (30)	16 (33)	17 (32)
Set E...	1	2	3	4	5	6	7	8	9	10	11	12
Alternative 1	13 (11)	6 (1)	9 (1)	6	**120** (1)	6	12	*33* (1)	**68**	*32*	7	*32*
2	— (1)	1 (1)	23	**115**	11	21	**37**	16	1	19 (1)	10 (1)	*32*
3	13 (3)	*40* (3)	9 (2)	*30* (3)	3	5	23	*29*	7	4	**57**	12 (1)
4	3 (6)	2 (2)	— (1)	14	15	14	19	**17**	16	—	*27*	3
5	8 (2)	2 (2)	7	8	1	**93**	4	13	*29*	13	7	**21**
6	4	**154** (5)	5 (2)	16	3	8	*24*	7	19 (1)	**57**	1	11
7	**175** (3)	7 (1)	— (1)	7	*46* (2)	19	*52* (1)	*27*	6	6	17	17
8	1 (1)	3	**159** (3)	13 (1)	4	*24* (1)	6	6	10	22	19	6
Not attempted	19 (8)	21 (20)	*24* (25)	*27* (31)	*33* (32)	*46* (34)	*59* (34)	*88* (34)	*80* (34)	*83* (34)	*91* (34)	*102* (34)

every case. Where Bromley's subjects chose a wrong response other than the one favoured by our subjects the response in question is typically 'unintelligent', usually involving the choice of a shape the same as the one immediately to the left of or immediately above the vacant space. Results for both samples are shown in Table 1.

Bromley administered the test individually, with no time limit, and recorded his subjects' explanations of the choices they made. It might be that if our subjects' explanations were available they would seem less primitive, but if we consider only the relative difficulty of the matrices and the frequency with which particular wrong responses were selected it appears that the results of Bromley's sample are typical of a normal population. Clearly their choice of wrong responses was not a symptom of age or mental abnormality except in so far as either factor might have brought about an overall reduction in level of intelligence.

The fact that a large proportion of subjects of above average intelligence frequently agree in supposing that a particular incorrect response to one of the matrices is correct seems to suggest that Spearman's three noegenetic principles (cf. Spearman, 1927) may not exhaust the possible modes of intelligent response to the matrices. It is beside the point to argue that Spearman's theory 'does not require to account for "incorrect" performance'; such a view assumes that Spearman's theory is demonstrably valid as a description of intelligent behaviour. The fact that intelligent individuals (that is, individuals capable of achieving a high total score on the test) behave as if other modes of response were permissible ought to lead us to question the adequacy of Spearman's description generally, not merely its adequacy as a description of the modes of response of 'primitive' individuals, because these modes are equally characteristic of normal individuals of above average intelligence.

REFERENCES

BROMLEY, D. B. (1953). Primitive forms of response to the matrices test. *J. ment. Sci.* **99**, 374–93.

RAVEN, J. C. (1938). *Progressive Matrices Sets, A, B, C, D and E.* London: H. K. Lewis.

SPEARMAN, C. (1927). *The Nature of Intelligence and the Principles of Cognition.* London: Macmillan.

(*Manuscript received* 2 *September* 1965)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 301–306
Printed in Great Britain

PERCEPTUAL SENSITIZATION, RESPONSE SUPPRESSION, AND PSYCHOPATHOLOGY

BY T. M. CAINE
Claybury Hospital, Woodford Green, Essex

A sentence-building test is described which elicits (i) 'acting out hostility' versus 'neutral' responses; (ii) 'projected hostility' versus 'neutral'; (iii) 'guilt' versus 'neutral'; (iv) 'neutral' versus 'neutral' responses. The test is composed of sets of sixteen items each, covering each of these four areas. Under two conditions of administration, one of which facilitates response suppression, normals were found to suppress the emotionally toned material in all areas. The psychotics suppressed emotionally toned material, when the conditions of administration provided opportunity, in an area *not* related to their psychopathology (the 'acting out hostility' area). The implications for psychological testing situations are discussed.

Significant differences in response to 'sentence-building' tests eliciting aggressive, guilty, and neutral constructions, between various clinically designated groups, have been found by Watson, Pritzker & Madison (1955) and by Caine (1960, 1963). The differences between the groups emerged in an administration of the test maximizing the opportunities for response suppression. In a study of normal subjects and psychotic depressive patients, with regard to the number of guilty constructions verbalized under two types of administration of a sentence-building test, the results suggested that both groups were equally perceptually sensitized to the guilty material, but that the normal group tended to suppress this type of response more often than the psychotic depressive group when given the opportunity (Caine, 1963).

The present investigation is an extension of this latter study, which has been broadened to include both a wider range of subjects and other types of sentence-building material.

METHOD

The test

In the sentence-building test constructed for the present study, four sets of sixteen items each were devised. The sets involved the following choice of constructions: acting out hostility versus neutral, projected hostility versus neutral, guilt versus neutral, and neutral versus neutral (to provide a base-line reaction time). Choice words were equated for position in presentation and, as far as possible, for length and word frequency. Four examples from each set of sixteen items are given below:

acting out hostility/neutral

(*a*) alive find	(*b*) coax stab	(*c*) Shoot I'll	(*d*) kick the
them burn	you I'll	you cheer	feed dog

projected hostility/neutral

(*a*) they'd me	(*b*) tell they	(*c*) its plot	(*d*) they ahead
cheat find	lies time	a joke	evil plan

guilt/neutral

(*a*) courage no	(*b*) am rotten	(*c*) to blame	(*d*) pluck no
collars I've	sleepy I	I'm apply	I've piano

neutral/neutral

(*a*) see bird	(*b*) candle bring	(*c*) the victor	(*d*) their its
the hear	the bricks	waiter he's	goal gang

Procedure

The test was administered in two ways:

In method of administration I, subjects were instructed as follows: 'I am going to show you a series of cards on each of which are four common words. By leaving one word out, and by rearranging the other three, a meaningful phrase or sentence can be constructed. I want you to give me the first three word phrase or sentence you can.' The type of phrase or sentence verbalized was recorded as each card was presented. There was some emphasis on speed of response, since a stop watch was prominently displayed and the response time was noted down in an obvious manner. The subject's score for the acting out, projected hostility, and guilt sets of items was the total number of such phrases or sentences constructed.

In method of administration II, subjects were again told that they were to be shown a series of cards on each of which were four common words. This time, however, they were told: 'It is possible to make up *two* meaningful three word phrases or sentences out of the four words by using two of the words twice. I want you to give me both phrases or sentences.' The type of phrase or sentence verbalized *first* was noted down and the subject's score was the number of acting out hostility, projected hostility, or guilty constructions he verbalized first. The emphasis on reaction time was the same as for method I.

In each method three trial cards giving neutral/neutral constructions were first presented and subjects who were unable to grasp the principles of the test after a trial demonstration were dropped from the experiment. Four patients only were eliminated for failure to grasp the instructions.

Subjects

Two similar samples of ninety subjects each were employed. Each sample contained three subgroups consisting of thirty normal, thirty paranoid, and thirty pyschotic depressive subjects, with fifteen female and fifteen male subjects in each subgroup. One sample was given the sentence-building test under method of administration I, the other under method II.

Hypotheses

It was hypothesized from previous work that response suppression would operate most strongly under method I, in association with aggressive or guilty material having no profound or psychopathological significance for the subject (Caine, 1963). From the same study it was anticipated that all subjects would demonstrate perceptual sensitization for such socially 'taboo' material.

The following specific hypotheses were formulated:

(1) *Projected hostility plus guilt.* Because these sentence-building sets were constructed specifically to mirror psychopathological feelings of the psychotic group of paranoids and psychotic depressives, no response suppression was anticipated for the psychotic sample tested under method of administration I; scores for the psychotic sample under method I should be similar to scores for the psychotic sample under method II. In the normal samples, however, response suppression should operate under method I conditions but not under method II. It follows that the differences between administrations should be greater for the normal sample than for the psychotic.

(2) *Acting out hostility.* Because the 'acting out hostility' set of items is not associated in a psychopathological way with either the normal or the psychotic group, response suppression was expected to operate in both groups equally under the choice conditions of method of administration I, and also but less under method II. The scores in both groups should be similar, and the scores under method I should differ from the scores under method II.

(3) *Reaction times.* Perceptual sensitization should be supported by a quicker reaction time for the aggressive and guilty constructions, than for the neutral/neutral base-line set, under both administrative conditions.

RESULTS

The mean scores of the combined sets of projected hostility plus guilt, under the two types of test administration, are shown in Table 1(*a*) and those for the acting out hostility set in Table 1(*b*).

The effects of response suppression in the normal and psychotic groups, in relation to the type of stimulus material presented and the type of test administration, are shown in Table 2, hypothesis (1), and Table 3, hypothesis (2), using the method of factorial design in the analysis of variance (Snedecor, 1957).

The reaction time comparisons, hypothesis (3), between the aggressive and guilty

Table 1. *Mean scores on the sentence-building test* (*a*) *for the 'projected hostility plus guilt' and* (*b*) *for the 'acting out hostility' items of normal, paranoid and psychotic depressive subjects on two types of administration of the test* (15 *males and* 15 *females in each group*)

	(*a*) 'Projected hostility plus guilt' Administration		(*b*) 'Acting out hostility' Administration	
	Type I	Type II	Type I	Type II
Normal males	17·40	19·27	9·27	10·07
Normal females	16·53	21·00	8·13	9·93
Paranoid males	19·13	20·07	9·00	9·73
Paranoid females	19·27	18·87	9·40	9·73
Depressive males	19·80	19·87	9·20	9·93
Depressive females	20·00	19·87	7·80	9·20

Table 2. 3 × 2 × 2 *factorial analysis of variance to test the null hypothesis that normals and psychotics do not differ in response to 'projected hostility plus guilt' items of the sentence-building test administered under two conditions*

Source	D.F.	S.S.	M.S.	*F*	*P*
Administration (A)	1	57·80	57·80	4·66	—
Sex (S)	1	0	0	—	—
Normal/psychotic (N/P)	1	44·80	44·80	3·61	—
A × S	1	1·42	1·42	—	—
A × N/P	1	93·03	93·03	7·50	< 0·01
S × N/P	1	4·23	4·23	—	—
A × S × N/P	1	28·34	28·34	—	—
Depressive/paranoia (D/P)	1	9·08	9·08	—	—
D/P × A	1	0·68	0·68	—	—
D/P × S	1	3·01	3·01	—	—
D/P × A × S	1	2·41	2·41	—	—
Within groups	168	2083·47	12·40	—	—
Total	179	2328·27	—	—	—

Table 3. 3 × 2 × 2 *factorial analysis of variance to test the null hypothesis that normals and psychotics do not differ in response to 'acting out hostility' items of the sentence-building test administered under two conditions*

Source	D.F.	S.S.	M.S.	*F*	*P*
Administration (A)	1	42·05	42·05	7·87	< 0·01
Sex (S)	1	11·25	11·25	—	—
Normal/psychotic (N/P)	1	0·40	0·40	—	—
A × S	1	2·01	2·01	—	—
A × N/P	1	2·50	2·50	—	—
S × N/P	1	0·40	0·40	—	—
A × S × N/P	1	1·88	1·88	—	—
Depressive/paranoia (D/P)	1	5·63	5·63	—	—
D/P × A	1	2·13	2·13	—	—
D/P × S	1	12·03	12·03	—	—
D/P × A × S	1	2·13	2·13	—	—
Within groups	168	897·13	5·34	—	—
Total	179	979·54	—	—	—

constructions and the neutral/neutral base-line constructions, are shown in Table 4, using greater or less than the base-line reaction time as the basis of comparison.

In Table 1 (*a*) the means of the normals under method of administration I are lower than the normals under method II and there is little difference in the psychotics between either administration. The analysis of variance of these data (Table 2) confirms these findings in that the difference between methods of administration is greater in the normal sample than in the psychotic. Hypothesis (1) is thus supported.

In Table 1 (*b*) all the groups have similar means under method of administration I, with the possible exception of depressive females. There is a general rise in score in all groups under method II. This is confirmed in Table 3. Only the comparison for methods of administration yields a significant F value. The prediction made in hypothesis (2) is thus sustained.

Table 4. *A sign test of the null hypothesis that emotionally toned response reaction times do not differ from base-line neutral response times*

	Frequencies			
	Greater than base-line	Less than base-line	P	
Normals				
Acting out hostility				
Admin. type I	7	23	8·533	< 0·01
Admin. type II	7	23	8·533	< 0·01
Projected hostility				
Admin. type I	10	20	3·333	< 0·06
Admin. type II	8	22	6·333	< 0·05
Guilt				
Admin. type I	5	25	13·333	< 0·001
Admin. type II	3	27	19·200	< 0·001
Psychotics				
Acting out hostility				
Admin. type I	18	41	8·881	< 0·01
Admin. type II	22	38	4·267	< 0·05
Projected hostility				
Admin. type I	15	45	15·000	< 0·001
Admin. type II	22	38	4·267	< 0·05
Guilt				
Admin. type I	5	55	41·683	< 0·001
Admin. type II	5	54	40·695	< 0·001

The results of these analyses suggest that, when opportunity is provided for response suppression (as in method of administration I), the expression of material suggesting that the subject is overtly aggressive, persecuted or guilty is likely to be avoided in favour of a socially innocuous response, unless involvement with these feelings is present to a pathological degree as in the psychotic group.

It seems likely that purely perceptual differences between normals and psychotics would have emerged under method II, but no significant differences between the groups were found under this administration. A significant number of subjects scored above chance levels, however, suggesting that most subjects were sensitized to the emotionally loaded material, regardless of their diagnostic category (psychotic or normal). Table 4, showing a quicker average reaction time for the same material compared with neutral material, supports this finding. This makes good sense in ordinary perceptual experience, where an alarm signal would be expected to arouse appropriate physical or psychological defensive action for successful adjustment or, indeed, survival (Berlyne, 1957; Vernon, 1964).

DISCUSSION

The results support the contentions of Edwards (1957) that 'social desirability' is a powerful, relatively all-pervasive influence in normal subjects in determining the response to personality inventories, and extend his argument to the more covert testing level of the sentence-building test technique. Pertinent, too, is the argument of Wiggins (1962) that 'individuals who ascribe socially undesirable statements to themselves are deviant...'.

Although the evidence presented here supports Edwards, it is to be doubted if social desirability can be regarded as a personality trait. It seems more reasonable to regard a psychological testing session as essentially a social interaction in which the subject and the tester are the active participants. The testee is responding not only to the test material but also to the presence of the tester. He will respond in terms of his interpretation of the situation as a whole. One of the fundamentals of successful communication between persons is the definition of 'the other' and oneself as individuals. This the subject will do by evaluating the tester as he attempts to establish rapport and administer the test and by selecting, editing, and reporting the material presented to him in such a way as to reflect himself as a person.

In the present study the patients may well have regarded the psychologist as a potentially interested and helpful person, to whom they might safely reveal something of their problems and fears. A patient's communications, in the areas related to his psychopathological attitudes, may well have had a compulsive character about them. That something of the sort was going on in the testing sessions was indicated by the tendency, under method of administration I, of some subjects to apologize after making a hostile sentence construction, or to confirm or deny the truth of a construction as it related to themselves, often with an embarrassed laugh, sneer or deprecatory tone of voice. The implications for the traditional projective tests are obvious.

Differences between the two groups could thus be interpreted in terms of the 'cognitive attitude' theory developed by Klein (1958) and his colleagues, in which a subject's *report* of what he sees in a perceptual experiment is an active process of sifting and choosing among alternatives, 'which always narrows a subject's report of his perception'.

The evidence for selective report, even in acute psychotic patients (Table 4), supports the arguments of Lazarus (1954), Eriksen (1954) and Chodorkoff (1954) that, in order to evoke unusual reactions in perceptual defence studies involving the report of words, only stimulus words *known* to be emotionally disturbing to specific individuals should be juxtaposed to the neutral material.

REFERENCES

BERLYNE, D. E. (1957). Conflict and information-theory variables as determinants of human perceptual curiosity. *J. exp. Psychol.* **53**, 399.

CAINE, T. M. (1960). The expression of hostility and guilt in melancholic and paranoid women. *J. consult. Psychol.* **24**, 18–22.

CAINE, T. M. (1963). Response suppression in normal and psychotic groups. *Psychol. Rep.* **12**, 942.

CHODORKOFF, B. (1954). Self perception, perceptual defense and adjustment. *J. abnorm. soc. Psychol.* **49**, 508–12.

EDWARDS, A. L. (1957). *The Social Desirability Variable in Personality Assessment and Research.* New York: Dryden.

ERIKSEN, C. W. (1954). The case for perceptual defense. *Psychol. Rev.* **61**, 175–81.

KLEIN, G. S. (1958). Cognitive control and motivation. In *Assessment of Human Motives* (ed. G. Lindzey). New York: Rinehart.

LAZARUS, R. S. (1954). Is there a mechanism of perceptual defence? A reply to Postman, Bronson, and Gropper. *J. abnorm. soc. Psychol.* **49**, 396–8.

SNEDECOR, G. W. (1957). *Statistical Methods.* Ames, Iowa: Iowa University Press.

VERNON, M. D. (1964). Attention and visual perception. In *Readings in Psychology* (ed. J. Cohen). London: Allen and Unwin.

WATSON, R. E., PRITZKER, L. & MADISON, P. (1955). Hostility in neurotics and normals. *J. abnorm. soc. Psychol.* **50**, 36–40.

WIGGINS, J. S. (1962). Definitions of social desirability and acquiescence in personality inventories. In *Measurement in personality and cognition* (ed. S. Messick and J. Ross). New York and London: Wiley.

(*Manuscript received* 31 *August* 1965)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 307–318
Printed in Great Britain

THE SIZE-CONSTANCY THEORY OF VISUAL ILLUSIONS

BY W. H. N. HOTOPF
London School of Economics and Political Science, University of London

This article criticizes the view that visual illusions, conceived as flat projections of typical views of objects lying in three-dimensional space, result from inappropriate constancy scaling. It points out that visual illusions are multiply determined and that the theory in some cases leads to contradictions. It suggests that perspective interpretations of visual illusions have overemphasized the influence of geometrical environments. Finally it criticizes attempts to distinguish between two types of scaling responsible for constancy.

I. INTRODUCTION

In recent times the perspective theories of visual illusions of Volkmann, Thiéry and others have been revived by Tausch (1954), von Holst (1957) and Gregory (1963). What has characterized their revival has been the further theory that the visual illusions are due to the faulty working of a correction mechanism that normally operates to produce size constancy. For Tausch this is the sole mechanism, for von Holst and Gregory, it is subsidiary, making a relatively minor contribution to constancy. Tausch's article is full of ingenious demonstrations showing how different visual illusions may be given perspective interpretations. His interpretations are somewhat *ad hoc*, however, and he made no attempt to relate his different interpretations one to another. On the other hand, Gregory's theory, which is more wide-ranging than that of von Holst, has been put forward in terms of a general principle and should therefore be easier to evaluate. For that reason the present article concentrates mainly on Gregory's theory. The reader should be warned, however, that though Gregory's theory has been stated in a variety of publications (e.g. Gregory 1963, 1966*b*), and has excited a great deal of interest, its statement is informal. The theory has not yet been published in any psychological journal, except by way of reply to a criticism (Gregory 1966*a*), and it should probably be regarded as still in the process of development.

II. STATEMENT OF THE THEORY

Gregory (1963) attributed size constancy to two different kinds of scaling, primary and secondary. The latter 'is set simply by apparent distance'. Errors in apparent distance produce corresponding size errors, of which 'the well-known moon illusion may be an example'. Primary scaling, on the other hand, 'is set by perspective and other features normally associated with distance'. It can operate without any *sense* of distance being present at all. This is where the visual illusions come in. According to the theory, 'The illusion figures may be thought of as flat projections of typical views of objects lying in three-dimensional space.' They therefore contain 'some features normally associated with distance', but we do not have an impression of distance because we see the illusion figures as drawn on a plane surface. Nevertheless, these features bring about corresponding size distortions in the other parts of the figures, such that 'the parts of the figures corresponding to distant objects are expanded and the parts corresponding to nearer objects are reduced'. The formulation implies that all visual illusions are composed of two parts. For the purposes of exposition in this paper, these parts are labelled the *distorting* and the *distorted parts*, though without the implication that the distortion is due to any feature associated with distance.

It is important to emphasize the distinction between primary and secondary scaling because recent criticisms, such as those by Brown & Houssiadas (1965) and by Hamilton (1966), have been based upon mistaken beliefs that primary scaling depends upon perception of distance or alone determines the amount of constancy. Although Gregory has been led to reject their criticisms, as a consequence, it is to be doubted if he has in fact demonstrated the independent existence of the two kinds of scaling. Further, the theory generates some contradictions and fails to explain certain features of the illusions. These criticisms are developed in relation to the Müller–Lyer illusion, to the so-called 'angle illusions', and to effects Gregory has obtained by presenting two-dimensional luminous models in the dark.

III. Criticisms of the theory

The Müller–Lyer illusion

Gregory derives the Müller–Lyer illusion in its vertical form from our perception of the corners of cubes viewed from outside and from inside. He regards the shorter-appearing line as a schematic representation of the vertical edge of a cube, extending above and below our line of vision, the arrows being the top and bottom sides of the cube receding from us; and the longer appearing line as, for example, the corner of a room, the arrows representing the intersections of the wall with the ceiling and the floor. Amongst other illustrations, Tausch (1954) compared the horizontal version of the longer line to the horizon-line of a road seen stretching before us, with the roof- and ground-lines of terraced houses on either side forming the arrow-heads—in other words, to a vista effect. Looking down a long narrow corridor would be another example of the same thing.

The evidence that Gregory has put forward for his theory includes the two following points. The first is the finding of Segall, Campbell & Herskovits (1963) that, as Gregory (1963) puts it, 'in primitive races living in houses without corners the geometrical illusions are reduced'. This is regarded as support for the theory because Gregory argues that primary scaling is learned (Gregory, 1963). The second point is Gregory's experimental finding (Gregory, 1966*b*) that there is a high correlation between degree of illusion and subject's judgements of how far back or how far forward the arrow-lines stretch, when the Müller–Lyer illusion is presented as a luminous figure in the dark and the narrowness of the arrow-heads is varied. What is found is that within limits the narrower the arrow-heads in either figure, the greater the illusion and the further forward or back the arrow-lines are seen as stretching.

Two difficulties for this theory are now discussed. First it is well known that the longer the arrow-lines relative to the connecting line, the weaker the illusion. Yet, as in the vista effect, depth should be most strongly experienced in these circumstances. The usual explanation is in terms of contrast, though Tausch (1954) gave a perspective interpretation of the 'contrast illusions', which Kristof (1961) has applied to the

Müller–Lyer. But whichever explanation is adopted, it can no longer be maintained that the whole of the illusion can be regarded as determined from a single viewing position.

The second difficulty for the theory as applied to the Müller–Lyer illusion lies in the circumstance that the narrowness of the arrow-heads, and therefore the degree of the illusion, depends upon height. For a given height, the narrowness of the arrow-head varies *inversely* with the distance from the observer. The farther away we move from a cupboard, for instance, the more the angles formed by the sides with the vertical edge facing us will flatten out. This entails the rather unlikely consequence that in Gregory's experiment referred to above (Gregory, 1966*b*), every time the angle is changed the subjects should perceive that the vertical line has changed in length or that their distance from it has changed. Further, the illusion may be at its strongest with an angle narrower than could be projected on the retina from a corner of 90°, as Mr Weinreich (personal communication) has pointed out. In Gregory's experiment, for example, the illusion with outward-going arrows is at its strongest when the angle enclosed by the arrow is as little as 60°. A more likely interpretation, that is still in accord with the perspective theory, would be that the illusion is determined by our experience with wedges, doors open at an angle and so on, rather than by 'typical views' such as corners in houses.

Angle illusions

The view has been widely held since the time of Wundt that illusions like those of Zöllner, Poggendorf, Hering and Wundt, and others involving distortions of circles and rectangles by transverse lines, are due to underestimating obtuse and over-estimating acute angles. If this is what we do the illusions would follow, but little evidence has been produced for this view, apart from the illusions themselves. It is safer simply to say that these illusions all manifest a tendency for the angles of intersection of the distorting with the distorted parts to appear nearer to right angles than they are.

This 'regression to right angles' tendency in the angle illusions is worth noting since it helps in the construction of modified figures and may have theoretical significance. It can of course be given a perspective interpretation: the intersecting lines can be interpreted as perspective drawings of right-angle intersections, as Sanford (1897), for example, recognized. But this interpretation is not the same as Gregory's, which is in terms of size. Whether a perspective interpretation is invoked or not, the description in terms of 'regression to right angles' is consistent with the illusions in question and Gregory's is not. This can be demonstrated in four cases.

(*a*) In the Zöllner illusion, the parallel lines appear to converge or diverge in accordance with 'regression to right angles'. Now this convergence or divergence could be taken as lines receding or coming nearer us. The perspective indications for this are said to be the transversals. But as the lines get further away, the angle of the transversals should change and they should come closer to one another. Since this does not happen, the transversals are not providing the perspective cues, i.e. the gradation, that primary scaling demands. The illusion occurs indeed even when the angles of the transversals alternate between 30° and 40°, as in Fig. 1. The parallel lines in Fig. 1 are at an angle of 45° to the page rather than in either of the main directions. This orientation increases the illusion in the case of the Zöllner figure, perhaps because lines going in one of the main directions are more resistant to distortion. But whatever is responsible is an additional factor to be taken into consideration in accounting for the illusion. Like the 'contrast' effect mentioned above it suggests that illusions are multiply determined and consequently that any theory that may give rise to quantitative predictions will be more difficult to confirm.

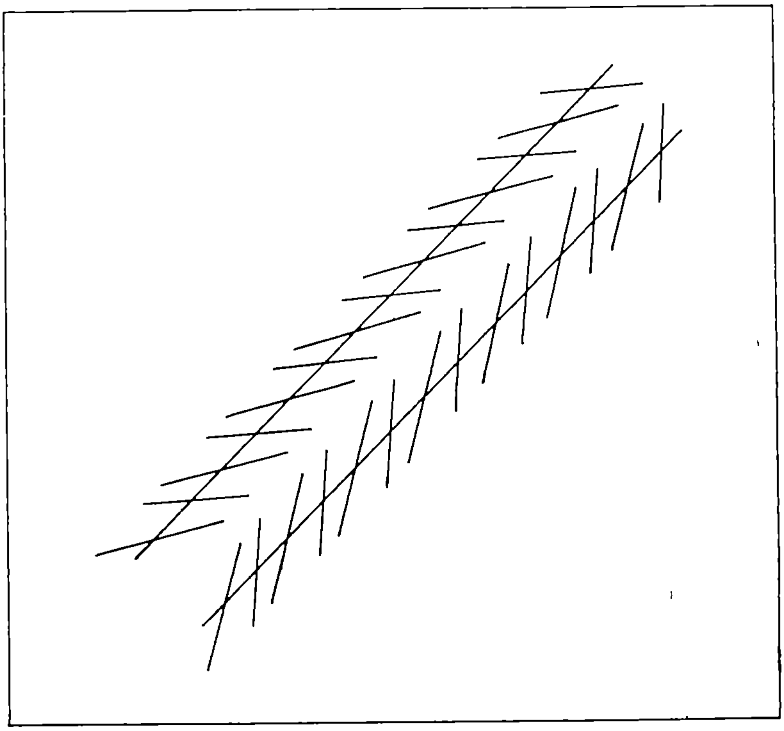
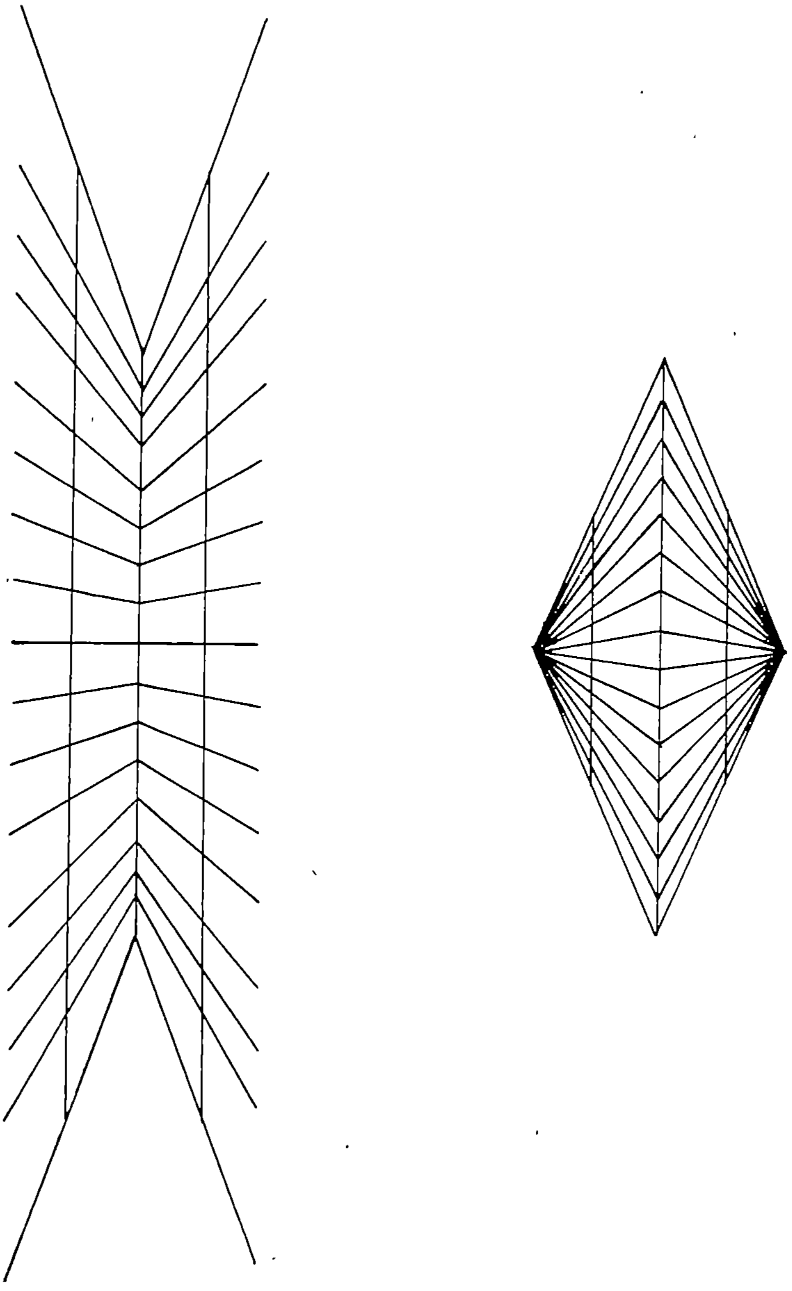


Fig. 1. The Zöllner illusion still operates with transversals alternating between 30° and 40°, though these would not be likely indicators of perspective. The 30° transversals are distorted, as well as the parallels.

(*b*) It is possible to combine the Müller-Lyer figure with Hering's and Wundt's figures, if the former is slightly modified (Fig. 2). It is clear that *both* pairs of illusions work. In Fig. 2*b* (Wundt's figure and the shorter Müller-Lyer) the two illusions can be explained by Gregory's theory. But the same explanation cannot be used for Fig. 2*a* (Hering's figure and the longer Müller-Lyer). If the vertical line in the centre of the figure looks longer than its counterpart in Fig. 2*b* because it is seen as farthest from instead of nearest to the observer, the observer's distance from the parallels must increase as they rise above or below eye-level. By Gregory's principle they should therefore seem wider apart as in Fig. 2*b*; but they are seen as nearer.

(*c*) There is an inconsistency in the explanation of the old illusion variously called Orbison's circle (by Gregory, 1963) and Ehrenstein's illusion (by Tausch, 1954). Adapting the distorting part of von Bezold's figure, I have drawn the circle on it so as to make the perspective effect more definite (Fig. 3). Von Bezold's figure is one of two alternative ways of presenting the tunnel or corridor illusion, where rectangles, cylinders or a boy and a man, though of equal size, appear different in size according to their position. Distance is suggested by the radiating lines. The distortion is



(*a*) (*b*)

Fig. 2. Combining the Muller–Lyer illusion with (*a*) a modified Hering figure and (*b*) the Wundt figure results in two pairs of combined illusions. Gregory's theory cannot explain (*a*) without self-contradiction.

consistent with 'regression to right angles' because, if the sides of the circle were bent to approach a right-angled intersection, the part of the circle nearest the point of origin of the radiating lines would be flattened. On Gregory's interpretation, on the other hand, this flattening is a sign of the circle being enlarged at its most distant part. But this argument considers size constancy without shape constancy. If we project a circular after-image on a receding plane what we also and more noticeably see is an elongation of the circle parallel to the line of vision, making it look like an ellipse. This may perhaps be seen to a slight degree on the other form of the corridor illusion (Fig. 4), where distance is suggested by vertical lines becoming closer together as in Gibson's three-cylinder illusion. This effect depends upon locating the circle on the wall; even if it cannot be detected, it is clear that there is no Orbison's circle effect as in the von Bezold figure, though the illusion is well marked in the case of the cylinders. What distinguishes the two settings of course is the absence or presence of the transversals that give the effect of 'regression to right angles'.

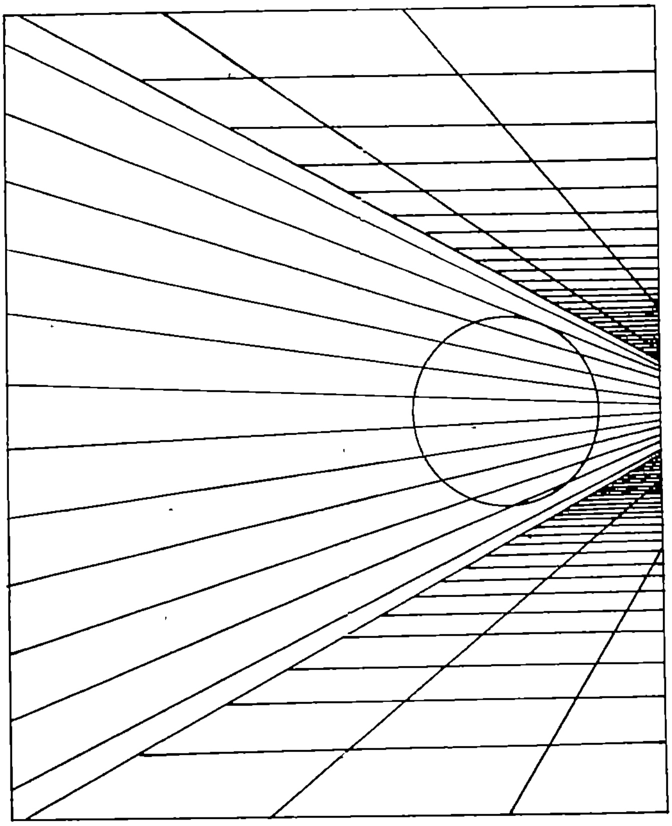


Fig. 3. A circle drawn on the background of a modified von Bezold's three-rectangle figure is not given the correct perspective elongation.

(*d*) Gregory has not indicated how his theory might account for the Poggendorf illusion, hence Tausch's explanation must be considered. Tausch (1954) regarded it as an instance of the tendency to overestimate the lengths of the arms of acute angles and to underestimate those of obtuse angles. This tendency he explained as due to

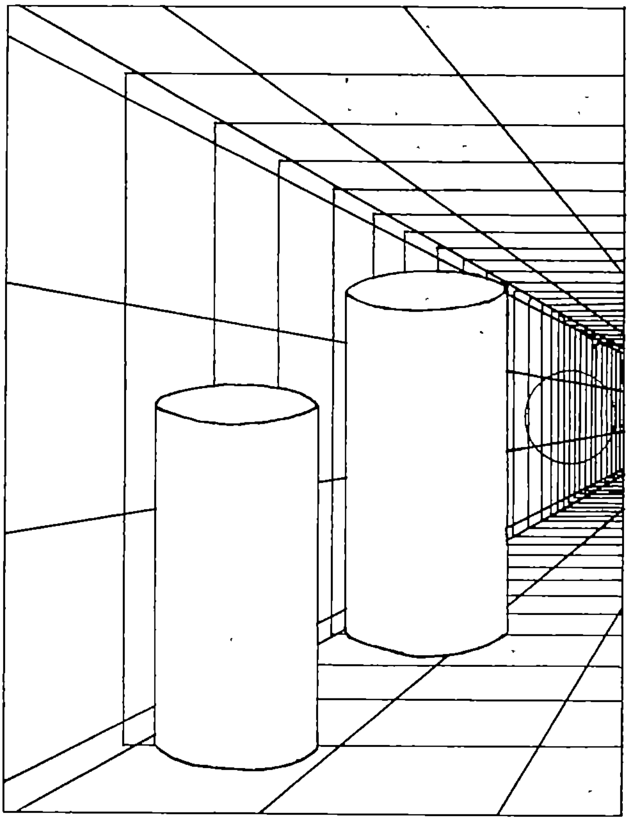
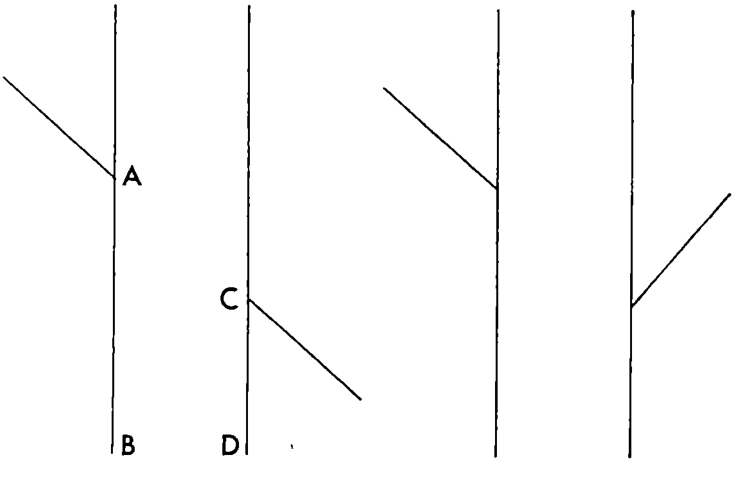


Fig. 4. A circle drawn on the background of Gibson's three-cylinder figure is given the correct perspective elongation.



(*a*) (*b*)

Fig. 5. The Poggendorf illusion remains though, according to Tausch's theory, changing the angle on the right-hand side from an acute to an obtuse one should abolish it.

the association of obtuse angles with the far corners of a receding rectangle, and of acute angles with the near ones. (Similarly Thiéry (1896) regarded the Müller–Lyer as deducible from the trapezoid illusions.) Applied to the Poggendorf figure, Fig. 5*a*, the side *AB* of the rectangle below the left-hand angle (an obtuse one) should appear longer than it is, whilst *CD* on the right side below the angle, which is of course acute, should appear shorter. Thus the right-hand angle should appear displaced downward, as is the case in the illusion. If the angle on the right side is reversed, however, as in Fig. 5*b*, the illusion is not cancelled out, but still remains, as has long been recognized. Further, as Burmester (1896) showed, the extent of the illusion depends upon the width of the rectangle. Tausch's theory cannot explain this, yet both this effect and the effects in Fig. 5 are consistent with 'regression to right angles'.

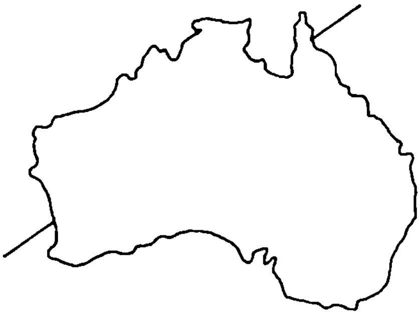


Fig. 6. The Poggendorf illusion does not depend upon having a regular figure for the indication of perspective.

The visual illusions are often referred to as the 'geometrical illusions', and since it seems likely that this term has encouraged the belief that they derive from a geometrical environment such as ours, it is worth pointing out that their geometrical character is merely a convenient convention. It is possible, for example, to produce the Zöllner illusion with the transversals roughly drawn in freehand. The Poggendorf is an illusion which Green & Hoyle (1963), for example, derive from perceptions of rectangular walls of rooms, but the distorting part need not be at all regular. The outline form of the continent of Australia is irregular enough, yet as Fig. 6 shows, it yields the Poggendorf illusion as effectively as a rectangle. Again, though the distorted part needs to be a regular or simple figure for an illusion to be demonstrated easily, there is no reason why converging as opposed to parallel lines, drawn only approximately straight, should not also be distorted. Visual illusions are not to be regarded as a limited set of curiosities, but rather as convenient indicators of distortions to be expected in ordinary vision, even if the perspective interpretation is not accepted as the whole explanation. This conception is extended further if it is accepted that a visual illusion can be subthreshold. Parallel lines, for instance, may be taken as a device for magnifying a deflexion of a single line by summing it with the contrasting deflexion of another line so that the deflexions will be perceived.

So far, for purposes of exposition, the visual illusions have been referred to as composed of two parts, a distorted and a distorting part. But, at least as far as the angle illusions are concerned, this view is mistaken. The distorting part itself is also distorted. For example, the Zöllner figure, as originally drawn with thick parallel lines and transversals, revealed the Poggendorf illusion, or 'nonius displacement' as it was called (Boring, 1942), in the downward displacement of the lower part of the transversals. The Poggendorf effect can also be seen in the Hering figure in Fig. 7. In the Zöllner illusion drawn with thin lines, though the nonius displacement cannot be observed (by hypothesis, below threshold), a tendency may nevertheless be detected for the ends of the transversals to bend in towards the line they transect. This effect is magnified so as to be unmistakable in the transversals at 30° in Fig. 1, particularly if the figure is turned through 45° so that the parallel lines are observed as verticals. It is difficult to say whether or not these additional distortions are ultimately reconcilable with 'regression to right angles', or indeed what their explanation may be. But, whatever the answer, they are difficult to reconcile with a conception that all visual illusions are explicable in terms of perspective effects from single viewing positions.

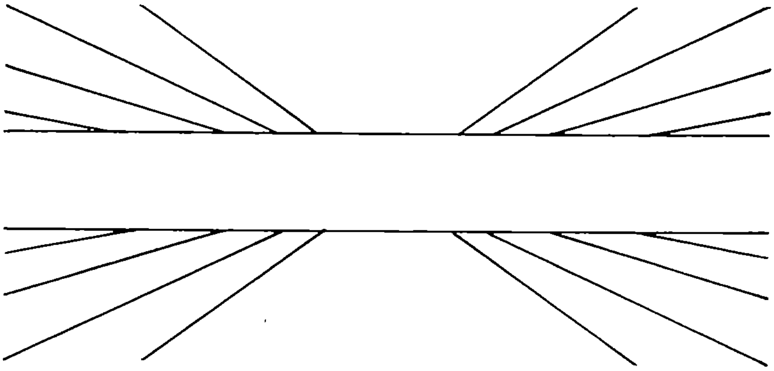


Fig. 7. This combination of the Poggendorf with the Hering figure shows that both illusions may work at once so that both parts of the figure are distorted.

IV. EFFECTS WITH LUMINOUS MODELS

Presenting luminous models of figures in the dark has been important for Gregory's theory, in two ways. First, the theory maintains that visual illusions are not seen as figures in depth when perspective cues are countermanded by others which indicate a plane surface; hence removing such indications should cause the figures to be seen as three-dimensional. Using luminous figures in the dark and viewing them monocularly enables the cues for plane surfaces to be removed. Gregory (1963) claimed that the expected effect is achieved under these conditions. But Green & Hoyle (1963) did not obtain the same result for the Poggendorf illusion and the writer did not obtain it in an experiment with the Müller–Lyer illusion. Luminous models of the Müller–Lyer figures were constructed; the stems were 15 in. and the arrows $5\frac{1}{2}$ in. long; both stems and arrows were 0·35 in. wide; the arrows formed angles of 65° and 115° respectively, with the stems in the shorter and the longer appearing figures. These figures were presented in the dark to twenty-five subjects who viewed them monocularly at a distance of 10 ft. None of these subjects saw both figures exactly as Gregory's theory would predict; sixteen subjects saw both figures as completely flat, and three saw some of the arrows as pointing in the wrong direction. It is, of course, perfectly possible for subjects to see the luminous Müller–Lyer figure as three-dimensional, but it is also possible for them to see a drawing of the illusion as a projection of a three-dimensional figure. Moreover, in view of the considerations advanced above in the section on the Müller–Lyer illusion, it is difficult to imagine Gregory's subjects both experiencing constancy and 'seeing' corners of 90°, when the connecting line in his experiment was only 10 cm long and was viewed from a distance of 50 cm (Gregory, 1966*b*). On the other hand, the situation is not dissimilar from one of judging distance in a perspective drawing.

The second way in which luminous models have been important is that Gregory has depended upon them to demonstrate the independent action of secondary scaling. Gregory (1966*b*) maintained that the sides of the Necker cube that change from near to far when it is reversed do not change in size in accordance with their change in distance. However, there *is* an appropriate size change when the cube is viewed as a two-dimensional luminous model. Therefore, he argued, the change in size, demonstrating constancy, is due to the perception of distance alone. If it had been due to perspective the Necker cube should have changed in size like a visual illusion when seen as a drawing on paper. The reasoning however is odd because if the Müller–Lyer figures are parts of cubes and are seen as they are because of the perspective indications, this must also be the case for Necker cube, which has *more* of these cues. The fact that they have alternative interpretations does not lessen their status as cues.

Gregory (1966*b*) by his further arguments in fact admits this point without seeming to be aware of the inconsistency. He maintains that a straight line drawn through a corner over two sides of the cube, as in Fig. 8, seems bent in accordance with one perspective interpretation of the cube, and does not change as it should when the perspective is reversed. He explains this in terms of his concept of a 'typical' perspective or viewing position—in this case, seeing the lower left face of the cube as to the fore. It is this in his view which determines the bend; and this primary scaling, aided presumably by cues from the surface of the paper, resists any secondary scaling influences when the perspective is reversed and we see from a 'non-typical' viewing position. On the other hand, he claims that when the cube is presented as a luminous, two-dimensional model in the dark, the bend in the line does change according to perspective position. The difficulty about this explanation is that the far side of the cube should look larger if we have primary scaling from the typical perspective position, and it should continue to look larger when the perspective is reversed and it is seen as the near side.

It may be noted in this connexion that Gregory's view that the far side of the Necker cube looks the same size as the near side for either perspective is not universally accepted. Sanford (1897), for example, in his treatment of visual illusions held that there was such a size change. It is, however, difficult to see with the ordinary Necker cube. On the assumption that the illusion is subthreshold, increasing the distance between the near and far sides of the cube should magnify it sufficiently to become detectable. Fig. 9 illustrates this point. This figure with sides 2 in. long was presented at eye level, at a distance of from 3½ to 4 ft. to thirty-two subjects.

They were asked to say for each perspective in turn whether the near side looked the same size, or whether it looked smaller or larger than the far side. Having done that, they were asked to say for each perspective whether the lines joining the near to the far sides looked parallel, or whether they converged or diverged. Thus they were given four opportunities of indicating a change in size in accordance with perspective. Only two subjects saw the sides as the same and the lines as parallel for both perspectives, i.e. saw it in the way claimed by Gregory. Of the rest, twenty-three indicated a change in size consistent with the far side looking larger in answer to three of the four questions put to them. It is notable that comparison of the size of the sides from the 'typical' perspective position was the least sensitive indicator; it should have been the most sensitive on Gregory's view.

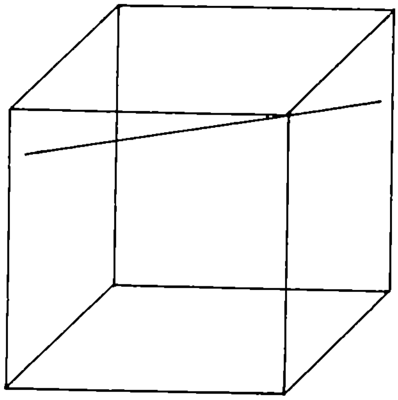


Fig. 8

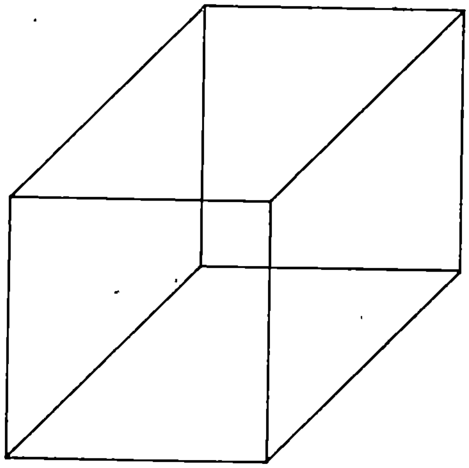


Fig. 9

Fig. 8. The failure of the apparent bend in the line to reverse itself when perspective is reversed is taken by Gregory as evidence that an illusion is determined by a typical viewing position, in this case one of looking from above downwards on the cube. He does not explain why the far side from this viewing position should not appear larger.

Fig. 9. An elongated Necker cube provides evidence that apparent size of near and distant sides changes appropriately with reversing perspective.

V. CONCLUSIONS

The foregoing arguments suggest the following conclusions.

1. There is no case on present evidence for distinguishing between primary and secondary scaling. It is indeed difficult to see how a form of scaling which was not itself due to 'perspective and other features associated with distance' (primary scaling) could be 'set simply by apparent distance' (secondary scaling).

2. That perspective might be responsible for *some* visual illusion is suggested by the evidence for an association between apparent distance and size in the case of the elongated Necker cube. In this case, however, the illusion is accompanied by an interpretation of the figure in terms of distance. In the case of the Müller–Lyer and the angle illusions, the illusion need not be accompanied by such an interpretation, i.e. the figure need not be seen as a flat projection of a typical view of an object lying in three-dimensional space for the illusion to occur.

3. Rather than being determined by a single factor, the illusions are determined by a number of factors, such as contrast, position with regard to the main directions in two-dimensional space, and so on. Even if these factors were considered inappropriately named and were given perspective interpretations instead as Tausch has attempted, the illusions still could not be explained in terms of single viewing positions.

4. The derivation of the Müller–Lyer illusion from experience of buildings, rooms and objects of furniture with right-angled corners is not well-founded. It is also suggested that the convention of presenting visual illusions as simple geometrical figures may have overdetermined the theory that they mainly derive from environments with similar properties.

The writer would like to express his gratitude to Miss Stephanie Hall of the Geography Department at the London School of Economics for drawing the figures used in this article.

References

Boring, E. G. (1942). *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton-Century.

Brown, L. B. & Houssiadas, L. (1965). The perception of illusions as a constancy phenomenon. *Brit. J. Psychol.* **56**, 135–41.

Burmester, E. (1896). Beitrag zur experimentellen Bestimmung geometrisch-optischer Täuschungen. *Z. Psychol.* **12**, 355–94.

Gibson, J. J. (1950). *The Perception of the Visual World*. Boston: Houghton Mifflin.

Green, R. T. & Hoyle, E. M. (1963). The Poggendorff illusion as a constancy phenomenon. *Nature, Lond.* **200**, 611–12.

Gregory, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature, Lond.* **199**, 678–80.

Gregory, R. L. (1966*a*). Comment on Dr Vernon Hamilton's paper. *Quart J. exp. Psychol.* **18**, 73–4.

Gregory, R. L. (1966*b*). Visual illusions. In *New Horizons in Psychology*, ed. B. M. Foss. Harmondsworth: Penguin Books.

Hamilton, V. (1966). Susceptibility to the Müller–Lyer illusion and its relationship to differences in size constancy. *Quart J. exp. Psychol.* **18**, 63–72.

von Holst, E. (1957). Aktive Leistungen der menschlichen Gesischtswahrnehmung. *Studium Generale* **10**, 231–43.

Kristof, W. (1961). Über die Einordnung geometrisch-optischer Täuschungen in die Gesetzmässigkeiten der visuellen Wahrnehmung. *Archiv. ges. Psychol.* **113**, 1–48, 127–50.

Sanford, E. C. (1897). *A Course in Experimental Psychology, Part I: Sensation and Perception.* London: Heath.

Segall, M. H., Campbell, D. T. & Herskovits, M. J. (1963). Cultural differences in the perception of geometric illusions. *Science* **139**, 769–71.

Tausch, R. (1954). Optische Täuschungen als artifizielle Effekte der Gestaltungsprozesse von Grössen und Formenkonstanz in der natürlichen Raumwahrnehmung. *Psychol. Forsch.* **24**, 299–348.

Thiéry, A. (1896). Ueber geometrisch-optische Tauschungen. *Phil. Stud.* **12**, 67–126.

(*Manuscript received* 26 *October* 1965)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 319–328
Printed in Great Britain

SIZE CONSTANCY AND INTELLIGENCE: A RE-EXAMINATION

BY VERNON HAMILTON
Reading University

This investigation was concerned with a re-examination of the usually accepted negative correlation between intelligence and size constancy. Size estimates of three standards were obtained from two groups of school-children of similar age but markedly different intelligence, and the incidence of over-constancy and under-constancy in different regions of a range of I.Q.'s was critically considered. The four major conclusions were that knowledge of the laws of perspective influences reduction in under-constancy, that under-constancy as well as over-constancy may occur with low as well as high I.Q.'s, that over-constancy appears to be the more usual and adaptive behaviour with higher cognitive development, and that significant changes away from under-constancy are associated with higher intellectual levels.

The most frequently held assumption concerning the relation between size-constancy judgements and intelligence is that they are negatively correlated, that the correlation is just barely significant at the 5% level, or that it just fails to reach statistical significance. The actual interpretation of this relation is that people of high intelligence tend to see an object in the distance as smaller or shorter than it actually is, and, in Thouless's terms, that less intelligent subjects show a greater tendency to perceive the 'real' size of objects (Thouless, 1932). Similar findings have been reported by Crookes (1957). A second frequently held assumption is that size constancy increases with age. The actual interpretation of this relation is that there is a reduction in the tendency to underestimate the size of an object in the distance—to make size judgements tending in the direction of retinal size—as children grow older and when children and adults are compared (Beyrl, 1926; Vernon, 1952; Zeigler & Leibowitz, 1957; Jenkin & Feallock, 1960). It has further been found (e.g. Thouless, 1932; Crookes, 1957) that what is termed under-constancy, i.e. under-estimating an object in the distance, continues to decrease during adulthood, and that over-constancy judgements are frequent in adults who, as far as can be inferred, are fairly or highly intelligent (Gilinsky, 1955; Carlson, 1960, 1962). A third, and clearly subsumed assumption in relation to the development of size constancy has been that it is affected by the individual reaching adequate levels of cognitive and specifically perceptual development in his experiential interaction with the environment (e.g. Vernon, 1952).

A perusal of the literature shows that it is virtually impossible to make direct comparisons between the findings of different investigators. A number of the classical studies contain methodological flaws, idiosyncrasies or methods of data transformation which tended to bias the nature, direction and interpretation of the experimental results in relation to theory (Vernon, 1952; Carlson, 1962), and there is some evidence that the term 'size constancy' itself has not been sufficiently carefully defined. Thouless's term of 'phenomenal regression to the real object' has now fallen somewhat into disfavour. Yet this term has a degree of descriptive directional precision which seems altogether lacking from terms such as: 'increases in size constancy', 'high constancy' or 'low constancy'. Furthermore, what is absent to the best of the writer's knowledge from all previous discussions in this field, is an attempt to clarify what is to be considered 'normal' or adaptive size constancy; whether, for example, full or nearly full accuracy of size estimation over distance is more appropriate, adaptive or useful to man than over-constancy with greater inaccuracy but indicating an awareness of perspective and of size/distance interaction, and an ability to counteract to primary sensory cues. It is known that horizontal angular separation and type of experimental instruction as well as their interpretation may affect the direction and extent of size-constancy estimates (Gilinsky, 1955; Joynson, 1958*a*, *b*; Carlson, 1960, 1962). It is still not clear, however, to what extent the relative position of standard and variable in the proximal or distal position affects constancy judgements in laboratory studies (Akishige, 1961) or how reliable the postulated overestimation of the standard is (Lambercier, 1946). Nor is it clear whether results obtained by means of the method of adjustment of a variable by subject or experimenter yield patterns of results which are not obtained or obtainable when the limiting or constant methods of psychophysical comparison are used.

The main reason for carrying out the present study was because findings from two previous size constancy experiments (Hamilton, 1963; Hamilton & Phillips, 1963) were apparently contradictory. They suggested that a negative correlation between size constancy and intelligence was for experimental purposes an inadequate description of the relation between these two variables which might bias the experimenter into formulating spurious conclusions. It was decided therefore to investigate the differential presence or absence of over-constancy and of under-constancy in samples of subjects of markedly different intelligence, of similar age, and with experimental instructions which included a recapitulation or description of perceived size/distance relations, and thus instructions which left no doubt in subjects that real size not phenomenal size matches were required. Gilinsky (1955) and Joynson (1958*a*, *b*) have already variously shown that differences in size estimation can be clearly due to differential interpretation of the task, and that analytic, retinal-type judgements may be made spontaneously with or without simultaneous fixation of two test objects by subjects of high intelligence. Nothing, however, can confidently be deduced from these findings that bears on the habitual or behavioural size-constancy responses of subjects with different degrees of intelligence, except that intelligent subjects are apparently able, when instructed, to make both analytic and objective size judgements.

METHOD

Apparatus. This has been described in detail elsewhere (Hamilton, 1963, 1966). Binocular constancy judgements were obtained in relation to three standard objects: a white square in the diagonal presentation (15·0 cm), a white rod (15·0 cm) and an Ace of Hearts playing card (8·90 cm) which were presented in this order. In the case of the square the variable was manually adjusted to size by the experimenter on instructions from subjects, employing a standard laboratory variable diamond. For the other two objects the variable was a white rod operated by subjects themselves by means of a handle. Each subject made four settings of the variable for each object, commencing adjustment alternately from minimal and maximal positions of the variable. There was a horizontal angular separation of 5° between standard and variable positions which were spatially arranged in proximal and distal order respectively. Viewing distances were 1 m and 3 m on a table painted matt black 3·10 m long, 0·70 m wide and 0·79 m high. A chin rest was used, and the data were obtained under full-cue conditions.

Instructions. As indicated earlier, the experimental aim was to maximize the probability that objective, 'real' size rather than apparent size judgements were made. To this purpose subjects were reminded that the same objects seem smaller when farther away in distance, and were given concrete examples of this relation in operation. They were instructed that the variable should be so adjusted that it would measure the same size as the standard if both were placed next to each other, or if both were measured. Instructions were continued until it appeared as certain as it could be that the requirements of the task had been grasped.

Subjects. Two samples of boys of secondary-school age were chosen. The average intelligence group came from the A, B and C streams of the first year of a Secondary Modern School, the low intelligence group was chosen from an appropriate age sample in a school for educationally subnormal children. The non-E.S.N. group was selected by its headmaster on the basis of their availability in terms of the school time-table, and the ages thus obtained were matched as closely as was possible with pupils from the E.S.N. school, excluding cases which in the opinion of the headmaster of the latter school would be unable to co-operate adequately because of sensory or physical defects or because there was evidence of a history of treatment for psychotic behaviour or of a history of epilepsy or cortical trauma. All subjects were normally sighted or were participating with corrected vision. These exclusive requirements prevented the selection of a fully age equated sample and the age difference between the two groups was in fact significant ($U = 114$; $P < 0{\cdot}02$, 2-tail). If age was as, or more, important than intelligence in the balance of under-constant and over-constant size estimates, then correlations with age should be of similar order to those with intelligence, and significant and consistent differences in over-constancy favouring the non-E.S.N. group should be more difficult to demonstrate. Conversely, if the expected group differences were nevertheless to materialize, they would gain somewhat in importance because they would be obtained in spite of age differences tending to vitiate them. It will be seen that the expected differences were found and it may merely be recorded that the correlations between constancy and age were close to zero.

The data reported here were obtained only as part of a wider study into perceptual/cognitive skills, and considerations of time prevented individual assessments of intelligence in conjunction with the assessment of size constancy. For the non-E.S.N. group the level of intelligence was obtained from subjects' scores on the Moray House Verbal Reasoning Test which they had all taken on the same day some nine months previously. For the E.S.N. group the latest individual I.Q. appearing in the School Medical Officer's files were used. These records are not always found to be reliable, and in this investigator's opinion some of the recorded I.Q. were lower than they might have been found to be on retest, while the scholastic performances of other children suggested that their true I.Q. might well have been somewhat lower than was indicated in the records. The absence of separate assessments of intelligence in conjunction with this study is undoubtedly a shortcoming, nevertheless it is clear that the two populations of subjects were of different intellectual calibres as shown by their placement in different schools. In the case of the non-E.S.N. group, subjects with the lowest Moray House Test I.Q. were in the B and C stream of their year. Table 1 shows the age and intelligence characteristics of the experimental groups.

Table 1. *Age and intelligence of the two groups of children*

		Secondary Modern School children ($n = 20$)	E.S.N. School children ($n = 20$)
Age	Mean	12·2	12·6
	S.D.	0·35	0·66
	Range	11 5–12·8	11·4–13 6
I.Q.	Mean	95·2	74·1
	S.D.	9·07	5·35
	Range	82–112	65–85

RESULTS

There was little to be gained from any arithmetical or algebraic transformation of the results, and therefore raw data in cm were the basis of all calculations. In the analysis of the present data the non-parametric Mann–Whitney U test was preferred wherever appropriate to the t test, and two-tail tests of significance were applied where a prediction of the direction of any difference could not validly be made. Under-constancy was operationally defined as any estimate greater by 5% or more than objective standard size, over-constancy was defined as any estimate smaller by 5% or more than standard size. Table 2 shows the means and measures of variability of the experimental groups for the three standard objects.

Table 2. *Means,* S.E. *and* S.D. *for the two* I.Q. *groups for each of the three test objects*

		Average I.Q.	Low I.Q.	Between-group differences
Diamond*	Means	15·06	16·18	P < 0·05
	S.E.	0·35	0·36	
	S.D.	1·51	1·55	
Rod*	Means	15·11	15·27	N.S.
	S.E.	0·33	0·51	
	S.D.	1·43	2·22	
Playing card†	Means	9·83	11·19	P < 0·025
	S.E.	0·30	0·46	
	S.D.	1·32	2·01	

* Veridical size, 15·00 cm; V.S. + 5%, 15·75 cm; V.S. − 5%, 14·25 cm.

† Veridical size, 8·90 cm; V.S. + 5%, 9·35 cm; V.S. − 5%, 8·46 cm.

When homogeneity of variance was tested in the two groups for each of the objects separately, none of the F ratios was significant. Subsequent discussion of group differences in size constancy estimates would make it legitimate, therefore, to prefer explanations in terms of true perceptual or cognitive performance rather than more *ad hoc* explanations in terms of haphazardness or inconsistency. With the considerable heterogeneity of intelligence, the product-moment correlations between size estimates and intelligence are appropriately higher than the correlations found by other investigators. They are for the diamond, rod and playing-card standards respectively: −0·446 (P < 0·01), −0·246 (N.S.) and −0·485 (P < 0·01) with $n = 40$. In the terms of the present investigation this means that subjects with low intelligence tend more generally to make under-constant size estimates and subjects with higher intelligence are more likely to make over-constant judgements. Subsequent analyses indicate, however, that this statement is only partially true.

Table 3 shows the distribution of constancy estimates for three standards, two groups and six I.Q. ranges. The data suggest that an inverse relation between intelligence and constancy is, to say the least, imperfect, but that for the non-E.S.N. group of average intelligence, considering the I.Q. range 81–110, there are some statistically significant trends linking increasing intelligence with reduction in under-constancy. If the two experimental groups are combined, as is legitimate since there are no significant differences in the I.Q. range 81–90 where the two groups overlap, they give the results shown in Fig. 1. The differential incidence of over- and under-

Table 3. *Relations between* I.Q. *and mean constancy for the three test objects*

		I.Q. range					
Standard	Group	61–70	71–80	81–90	91–100	101–110	111–120
Diamond	Av. I.Q.	·	·	15·97*	14·97*	13·60*	16·65
	Low I.Q.	16·55	16·06	15·76	·	·	·
Rod	Av. I.Q.	·	·	15·51	14·98	14·95	13·95
	Low I.Q.	16·35	15·09	13·14	·	·	·
Playing card	Av. I.Q.	·	·	10·60*	9·86	8·62*	10·33
	Low I.Q.	12·01	10·73	11·54	·	·	·
	n's	0	0	7	7	5	1
		6	12	2	0	0	0

* The differences between pairs of means in each row are significant ($P < 0·05$).

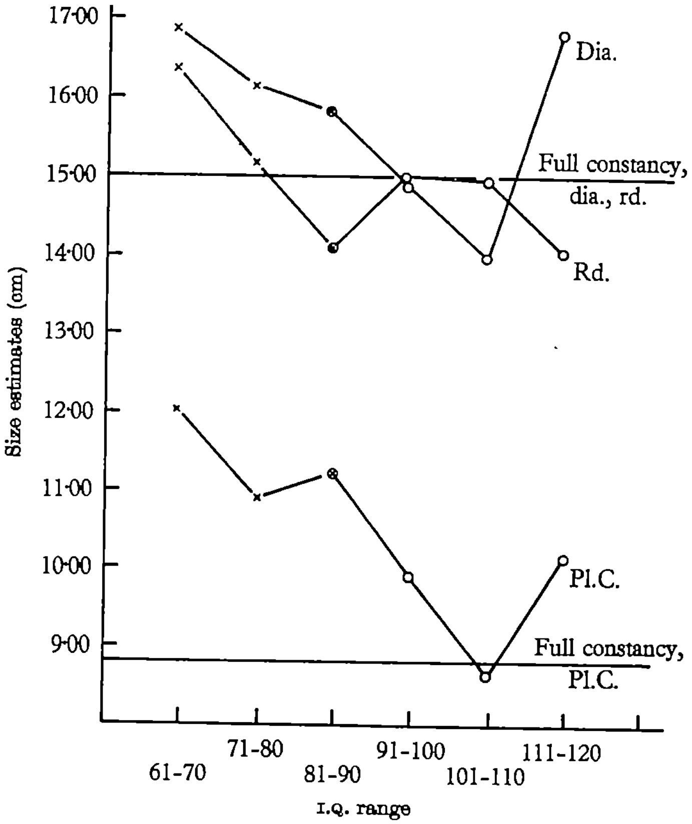


Fig. 1. Reduction of under-constancy with increasing I.Q. (×—× E.S.N.; O—O non-E.S.N.)

constancy in ranges of increasing I.Q. may be demonstrated more clearly by considering them separately and side by side as in Fig. 2. Combining responses to all three standards and calculating the percentage incidence of under-constant and over-constant responses, it appears that whereas the incidence of over-constant judgements increases smoothly and fairly steeply for both experimental groups, under-constant responses do not in fact decrease with higher intelligence in the E.S.N. sample, and they reappear with greater frequency in the highest intelligence group tested.

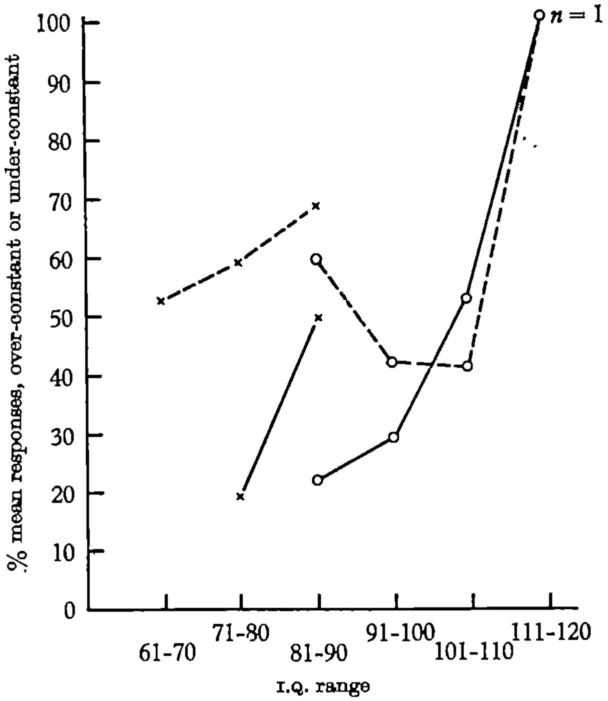
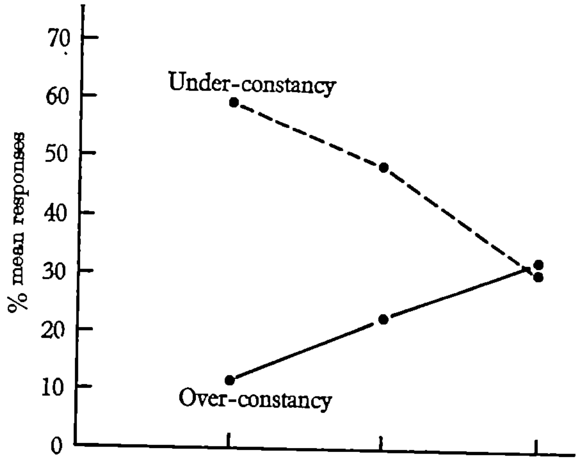


Fig. 2. Incidence of under-constancy and over-constancy in different I.Q. ranges, Over-constant: ×——×, E.S.N.; ○——○, non-E.S.N. Under-constant: ○- - -○, non-E.S.N.; ×- - -×, E.S.N.

While many studies involving children, adolescents and adults without behavioural abnormalities have been carried out, the distributions of over-constant and under-constant estimates have not been given separately in the past. The present differences in the distribution of under- and over-constant estimates between the two groups of different intelligence were not significant ($\chi^2 = 1{\cdot}6$), but it has already been shown that under-constancy occurs with diminishing frequency and over-constancy with increasing frequency with increases in intelligence. This investigation did not set out to test the *differential* inportance of age and intelligence in the development of size constancy. To do so adequately would require testing at frequent points of a fairly long age range. Even so there would be no guarantee that subjects in different age-groups are necessarily equated for those cognitive skills or operations, or for those perceptual experiences which play the major role in size constancy judgements.

Nevertheless, the present differences in constancy obtained from groups of similar age but different intelligence suggest that age may be important in the development of size constancy only in so far as it increases perceptual experience of size/distance relations and increases the probability that objective and valid intellectual/perceptual inferences about this relation can be made. It may be of added interest, therefore, to compare the present findings with results obtained from adults in a previous study (Hamilton, 1963) under identical conditions and with the same standards. These comparative data are plotted in Fig. 3. The plots show that over-constancy increases and under-constancy decreases into adult life and in the presence of good average to superior intelligence.



	E.S.N.	Non-E.S.N.	Adults
Mean age	12·6	12·2	34·5
Mean I.Q.	74	95	117
n	20	20	12

Fig. 3. Constancy and age.

DISCUSSION

The main purpose of the experiment was to determine the differential frequency of under-constancy and over-constancy in size estimates for different types of standard made by groups of clearly different intellectual capacity. The E.S.N. group was expected to produce a greater incidence of under-constant judgements as compared with the group of average intelligence. Rather surprisingly this was not found for the rod for which 25% of the responses of the E.S.N. group were over-constant estimates, and the mean values for this group were not greatly different from those for the non-E.S.N. group, although the variability was greater. The more complex stimulus objects produced significant differences between the groups, with the low-intelligence sample having less than half the number of over-constant responses of the more intelligent group. In general an examination of the three scatter diagrams shows that the negative correlation between size constancy and intelligence is primarily determined by a prevalence of under-constancy in the lower I.Q. ranges. While the frequency of over-constant estimates undoubtedly rises with greater intelligence, a sufficiently high number of under-constant and a sufficient paucity of over-constant estimates remains to depress the size of the correlation. The presence of tendencies in a group of relatively homogeneous intelligence both to overestimate and underestimate the size of objects in the distance, suggests that experiments in size constancy that have been concerned primarily with the influence of visual field factors, sensory characteristics and physical distance, may well require partial re-examination. For example, a study by Joynson, Newson & May (1965) has led these authors to suggest that a tendency to overestimate size in the distance (which they did not find to be a general phenomenon) may, in the case of very small objects at great distances, be a function of the very small angles which these objects subtend on the retina. Their own data tended to show that underestimation at a distance was less likely to occur when the angle subtended was smaller than between 1° and 2°, roughly the size of the fovea. In the present study, the angles subtended by variables set at 3 m to 150 cm and 8·90 cm were roughly 2·75° and 1·75° respectively. Over- and underestimation were present with both angular sizes, and rather remarkably the object subtending the smaller angle produced a greater average *underestimation* of size than the objects subtending the larger visual angle. It may be added that a visual angle/foveal size theory would have some difficulty in accounting for constancy differences varying with intelligence.

Nothing has been said as yet as to why there should be a relation between size constancy and intelligence, whatever its nature. While a theory of full covariance between perceived size and perceived distance can no longer be sanguinely accepted, there is little doubt that knowledge of the law of the visual angle, and of the law of perspective, and the ability to carry out perceptual inferences given maximal visual cues are of immediate and fundamental importance to a cognitive performance which, to be adaptive, must partially ignore primary sensory cues. It was predicted for the present study that the E.S.N. group of children would not only be of very significantly lower test intelligence, but that they would be significantly less aware of the parameters entering into perspective. Carlson (1962) had already found that an instructional bias towards perspective-orientated constancy responses leads to an even greater degree of over-constancy than objective size-orientated instructions. In a simple test situation, which was associated with the present investigation but cannot be discussed in detail, the author has also found that whereas full knowledge of perspective as operationally defined was present in 95 % of the non-E.S.N. group, no E.S.N. subject showed such a degree of awareness. The influence of this particular cognitive skill of knowing about and being able to apply the size-reducing aspects of perception in depth, linked with the presence or absence of conservation and seriation skills in Piaget's sense, was tested by calculating partial correlations. Holding these cognitive skills constant, the correlation between intelligence and size constancy in relation to the playing card was reduced to $-0{\cdot}331$, a value which is not significantly different from zero, and to $-0{\cdot}422$ for the diamond, a value which, though reduced, remains significantly different from zero at the 1 % level. The correlations between the sum of the cognitive skills and constancy were: $-0{\cdot}252$, $-0{\cdot}045$ and $-0{\cdot}380$ ($P < 0{\cdot}05$) for the three standards. Holding I.Q. constant the significant correlation was reduced to close to zero. While this evidence is not conclusive, it adds considerable support to the contention that in so far as intelligence has been shown to be related to size constancy, the relation may be partially mediated by the degree and nature of the development of fairly primitive cognitive skills. Deficiencies in this development are as likely to determine inaccurate inferences about the size of objects, and particularly of known objects seen at a distance, as they are likely to affect the development of a general skill of perception and of educational potential by virtue of primitive reasoning. It is in the nature of the design of currently used intelligence tests that these deficiencies may also be described by low I.Q.

The presence of *both* over-constancy and under-constancy in the higher or superior intelligence samples may well be explicable in broadly similar terms. The ability to make size judgements of reasonable accuracy despite discrepant retinal cues develops in everyone. Even subjects of low intelligence or young children make size estimates which are closer to objective than to retinal size. The skill evidently requires visual perceptual experience and is markedly underdeveloped in adults becoming sighted for the first time. Since size judgements of one kind or another are being continuously made in all kinds of adaptive behaviour, over-learning and thus over-constancy may be considered to be the more usual and more adaptive behaviour. At the adult level, a conscious attempt is frequently reported to compensate for the effects of distance if the task requires objective size estimates. This aim may result in a high incidence of over-constant judgements, because, though visually the distant object will in fact appear smaller, its apparent size will be justified by taking the distance effect into account. In the more intelligent respondent, under-constant estimates may nevertheless be made by taking size/distance inferences a stage further. These subjects may try to over-correct in the opposite direction for what they know they have to do in order to overcome the retinal cues. They know and infer what they have to do to allow for the distance effect, and then make an inference about their inference which may easily take their estimates back in the general direction of retinal size and thus of under-constancy. It does not seem implausible to suggest that this kind of self-correcting double inference is less likely in individuals of low I.Q.

The major implication seems to be that investigations of the processes associated with or determining size constancy responses would have to control not only for knowledge of visual angle and perspective, for intelligence and for age, but also for the presence of consistent and established response patterns in relation to objects of different complexities and meaning, viewed at a distance.

I am grateful to Dr D. H. Taylor for his assistance with the computing of some of the results, and to Mr W. A. Goodworth and Mr B. H. Tuggey, Headmasters of the schools supplying the subjects, who were most helpful in making research facilities available.

REFERENCES

AKISHIGE, Y. (1961). *Experimental Researches on the Structure of the Perceptual Space.* Kyushu Psychol. Stud. Fukuoka: Kyushu University.

BEYRL, F. (1926). Über die Grössenanffassung bei Kindern. *Z. Psychol.* **100**, 344–71.

CARLSON, V. R. (1960). Overestimation in size constancy judgments. *Amer. J. Psychol.* 73, 199–213.

CARLSON, V. R. (1962). Size constancy judgments and perceptual compromise. *J. exp. Psychol.* **63**, 68–73.

CROOKES, R. G. (1957). Size constancy and literalness in the Rorschach test. *Brit. J. med. Psychol.* **30**, 99–106.

GILINSKY, A. S. (1955). The effect of attitude upon the perception of size. *Amer. J. Psychol.* **68**, 173–192.

HAMILTON, V. (1963). Size constancy and cue responsiveness in psychosis. *Brit. J. Psychol.* **54,** 25–39.

HAMILTON, V. & PHILLIPS, J. P. N. (1963). Some further parameters of abnormal size constancy in schizophrenia. *Bull. Brit. Psychol. Soc.* **16,** no. 51. 21A–22A.

HAMILTON, V. (1966). Susceptibility to the Müller–Lyer illusion and its relationship to differences in size constancy. *Quart. J. exp. Psychol.* **18,** 63–72.

JENKIN, N. & FEALLOCK, S. M. (1960). Developmental and intellectual processes in size-distance judgment. *Amer. J. Psychol.* **73,** 268–73.

JOYNSON, R. B. (1958*a*). An experimental synthesis of the associationist and Gestalt accounts of the perception of size. Part I. *Quart. J. exp. Psychol.* **10,** 65–76.

JOYNSON, R. B. (1965*b*). An experimental synthesis of the associationist and Gestalt accounts of the peception of size. Part II. *Quart. J. exp. Psychol.* **10,** 142–54.

JOYNSON, R. B., NEWSON, L. J. & MAY, D. S. (1965). The limits of over-constancy. *Quart. J. exp. Psychol.* **17,** 209–16.

LAMBERCIER, M. (1946). La constance des grandeurs en comparaisons sériales. *Arch. Psychol.* **31,** 79–282

THOULESS, R. H. (1932). Individual differences in phenomenal regression. *Brit. J. Psychol.* **22,** 216–41.

VERNON, M. D. (1952). *A Further Study of Visual Perception.* Cambridge University Press.

ZEIGLER, H. P. & LEIBOWITZ, H. (1957). Apparent visual size as a function of distance for children and adults. *Amer. J. Psychol.* **70,** 106–9.

(*Manuscript received* 23 *September* 1965)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 329–333
Printed in Great Britain

THE EFFECT OF INFORMATIONAL FEEDBACK ON SIZE JUDGEMENTS OF SCHIZOPHRENIC PATIENTS

By R. STANNARD, G. SINGER AND R. OVER
Psychology Department, University of Sydney

This experiment studied visual size judgements of hospitalized schizophrenic patients and normal controls under two conditions of presentation. Sixteen patients and sixteen normal volunteers were required to match for size a graded series of disks at a distance of 14 ft 6 in. with a 4 in. standard at a distance of 5 ft 6 in. Both groups were tested under two conditions: verbal instructions, and verbal instructions with information feedback. The results showed a reduction in error for schizophrenics under the feedback condition. These findings are discussed with reference to the research methodology of procedures which are supposed to provide relative measures of the perceptual capacities of different diagnostic groups.

The present experiment was on the question of whether differences found in size judgements between normal and schizophrenic subjects represent differences in perceptual capacities (Hamilton, 1963; Lovinger, 1956; Rausch, 1952; Weckowicz, 1957). Garner, Hake & Eriksen (1956) have distinguished between perceptual and response functions. They have argued that a single response difference cannot be validly interpreted as an index of perceptual capacity; the extent to which differences are invariant with changes in non-sensory variables must be examined. In these terms differences in size judgements between normal and schizophrenic subjects may not indicate that the two groups respond differently because they perceive differently; it may be that the two groups perceive in a similar fashion but respond differently. This issue can be considered by examining the effects of non-sensory variables on judgements made by schizophrenics and normals.

Hunt and Cofer (1944) and Huston & Shakow (1948) have applied this type of argument to data on non-perceptual tasks. They have claimed that the performance deficits found for many tasks with schizophrenics may arise from motivational rather than cognitive deficiencies. Thus changes in performance are to be expected when the cognitive features are varied. Studies using aversive stimulation ('white noise') as motivation (Pascal & Swenson, 1952; Cavanaugh, 1958) have shown schizophrenic performance to be a function of level of motivation. Food has also been used to obtain effective control over the behaviour of hospitalized patients (Ayllon & Haughton, 1962). The results have been inconclusive when verbal encouragement or exhortation has been used (Stotsky, 1957; Topping & O'Connor, 1960).

Instructions are an important class of non-sensory variables that control size judgements (Gilinsky, 1955). Sets of words operate as instructions through a learning history in which the subject is informed of the consequences of different discriminations (Edwards, 1961). Some verbal instructions are ambiguous in that they produce modifications in judgements both between and within subjects (Joynson & Newson, 1962). Instructions can be made unambiguous by providing the subject with information about the accuracy of his judgements (Edwards, 1961). It is possible that differences in size judgements found between normals and schizophrenics may represent the differential meaning of instructions to the two groups rather than differences in perceptual capacity. If this were so differences in size judgements would not be expected when both groups are given unambiguous instructions. If differences, however, are independent of the instruction conditions, it can be concluded that they represent differences in perceptual capacities. The present experiment examines this issue.

METHOD

Subjects

A group of sixteen hospitalized schizophrenic patients and sixteen normal volunteers acted as subjects. The main criteria for the selection of schizophrenic patients as subjects were psychotic ideation or thought disorder, for example neologisms, thought blocking and pressure of thought, either at the time of testing or reported in the case history within the previous 2 months; adequate eyesight; and period of hospitalization less than 1 year. All subjects in the normal group had at least 20/20 vision with each eye. The mean age of the schizophrenic subjects was 28 years; of the normals 24 years. It was not possible to match the two groups in intelligence. The mean I.Q. for normals, prorated from scores on the Comprehension and Vocabulary subtests of the Wechsler Bellevue Intelligence Scale Form II, was 123·8; for schizophrenics 99·4.

Apparatus

The equipment consisted of the size judgement apparatus and apparatus to allow information feedback of results. The standard stimulus was a white circle, 4 in. in diameter, mounted on a screen 21 in. by 36 in. which was 5 ft 6 in. from the subject. The comparison stimulus, mounted on a screen of similar size at a distance of 14 ft 6 in. from the subject, was one of a series of eleven circles with diameters ranging from 3·375 to 4·625 in. in 0·125 in. steps. The stimuli were at right angles to each other.

There were three boxes on the table in front of the subject. On the middle box there was a button marked 'Same' and another marked 'Different'; the subject operated these buttons to signify his judgement. The other two boxes had glass panels, one marked 'Right' and the other 'Wrong'. These panels could be illuminated, by white and red light respectively, by means of switchgear and electronic timers out of sight of the subject. It was possible to pair buttons and lights so that, when the subject pressed one of the two buttons after the presentation of a particular comparison stimulus, one of the two lights came on for 1·5 sec. Silent switches were used to prevent the subject from using auditory cues.

Procedure

Half of each of the normal (N) and schizophrenic (S) groups were allocated to each of the two experimental conditions. Under one condition (V) subjects were instructed to judge whether the standard and comparison stimuli were the same or different in physical size on each presentation, and were given no information about the accuracy of their judgements. Under the other condition (IF) subjects were given the same verbal instructions but they were also told that they would be informed whether their judgements were correct or not. Thus there were four groups: normals under verbal instructions (NV), schizophrenics under verbal instructions (SV), normals under verbal instructions with information feedback (NIF), and schizophrenics under verbal instructions with information feedback (SIF). Information feedback was given by the 'Right' or 'Wrong' light panels. 'Same' and 'Right' were paired for the presentation of the 3·875, 4 and 4·125 in. circles of the comparison series, and 'Different' and 'Right' for the other eight circles.

With an informational feedback schedule it was considered necessary to have stimulus presentations arranged so as to lead to approximately equal number of 'Same' and 'Different' presentations since the predominance of one type of presentation could develop expectations which may differ for normal and schizophrenic subjects (Galanter, 1962). Thus, a trial was made up by each of the three 'Same' stimuli being presented three times and each of the eight 'Different' stimuli being presented once. The order of presentation within a trial varied over trials. Each subject was given ten trials (170 judgements) over two testing sessions.

RESULTS

Table 1 shows the percentage frequencies of 'Same' responses given by the four groups for the 'Different' comparison stimuli. The mean equality judgements of the four groups were accurate (4·03 *NV*, 4·01 *NIF*, 3·97 *SV*, 4·03 *SIF*) and do not differ significantly ($P > 0{\cdot}05$). In addition to the mean equality judgement, the variability of judgements around the mean was calculated for each subject. The means and S.D. of these variability measures are shown in Table 2. An analysis of variance indicated that the groups differ significantly ($F = 11{\cdot}50$, D.F. 3, 28, $P < 0{\cdot}01$) in intra-subject variability. Comparisons using Scheffé's procedure (Scheffé, 1953) show that the schizophrenic group given verbal instructions differs significantly ($P < 0{\cdot}01$) from the other three groups.*

Table 1. *Percentage frequencies of 'Same' responses for the experimental groups of normal (N) and schizophrenic (S) subjects, with (IF) and without (V) information feedback*

Diameter of target stimulus (in.)	*NV*	*NIF*	*SV*	*SIF*
4·625	—	—	15·0	—
4·5	2·5	6·3	17·5	6·3
4·375	20·0	16·3	28·8	25·0
4·25	57·5	33·8	36·3	63·8
4·125	81·0	76·5	55·8	78·5
4·0	88·0	91·0	66·5	86·0
3·875	69·8	77·5	67·5	77·5
3·75	32·5	45·0	58·8	42·5
3·635	8·8	7·5	36·3	8·8
3·5	3·8	—	23·8	6·3
3·375	—	—	12·5	—

Table 2. *Means and standard deviations of intra-subject variance scores based on* 170 *judgements by each subject*

	Group			
	Normal without information, *NV*	Schizophrenic without information, *SV*	Normal with information, *NIF*	Schizophrenic with information, *SIF*
Mean	0·020	0·045	0·020	0·023
S.D.	0·005	0·018	0·003	0·007

DISCUSSION

The results have shown that, although the mean equality settings of both schizophrenic and normal subjects were accurate, i.e. independent of instruction conditions, informational feedback led to a significant reduction in the intra-subject variability of schizophrenic judgements. The data from experimental studies that have been concerned with mean equality judgements for schizophrenics and normals are equivocal (Lovinger, 1956; Rausch, 1952; Weckowicz, 1957). Differences in mean equality judgements between schizophrenics and normals have generally been found only with reduction of viewing conditions (Lovinger, 1956), not under maximal conditions (Leibowitz & Pishkin, 1960). The maximal conditions used in this experiment showed no difference in the mean equality settings for the four groups but showed significantly larger intra-subject variability for the schizophrenics under verbal instructions. The difference in mean intra-subject variability indicates that with verbal instructions alone schizophrenics were giving 'Same' judgements somewhat indiscriminately. With knowledge of results, however, schizophrenic subjects came to use the 'Same' judgemental category in a similar manner to normal subjects.

---

* An analysis of variance of transformed variability scores (log variance) showed the same significance levels.

The design of the present experiment does not allow the different responses made by schizophrenic subjects under the two conditions to be attributed to motivational or to instructional aspects of the procedure. It could be argued that schizophrenics made inaccurate settings when verbal instructions alone were given because they were not motivated to respond accurately, and that the procedure whereby 'Right' was given to particular settings exerted a motivational control on discrimination. But it could also be argued that the payoff operates in an entirely instructional fashion by 'sharpening' the subjects' criterion of 'Same'. The latter viewpoint would require that the verbal instructions had different meanings for schizophrenic and normal subjects.

The experiment has demonstrated the need to employ converging operations (Garner, *et al.*, 1956) before interpreting judgemental differences in terms of differences in perceptual capacities. A difference in variability in size judgements of schizophrenic and normal subjects was not found under all instructional conditions. It is more profitable to speak of the greater variability of judgements found for schizophrenic subjects when verbal instructions alone were given as due to the differential control of non-sensory variables, rather than as a symptom of an underlying disturbed state. In the same way it may be more profitable to consider differences in mean judgements found between schizophrenics and normals under reduced viewing conditions in terms of instructional control rather than in terms of such vague processes as 'over-inclusion' (Perez, 1961) and 'poor contact with reality' (Lovinger, 1956). These terms suggest differences in perceptual capacities. Before such conclusions can be drawn, however, it would be necessary to employ converging operations to show that judgemental differences remain invariant when response variables are changed.

REFERENCES

AYLLON, T. & HAUGHTON, E. (1962). Control of the behaviour of schizophrenic patients by food. *J. exp. Anal. Behav.* **5**, 343–52.

CAVANAUGH, D. K. (1958). Improvement in the performance of schizophrenics on concept-formation tasks as a function of motivational change. *J. abnorm. soc. Psychol.* **57**, 8–12.

EDWARDS, W. (1961). Costs and pay-off are instructions. *Psychol. Rev.* **68**, 275–84.

GALANTER, E. (1962). Contemporary psychophysics. In *New Directions in Psychology*. New York: Holt, Rinehart and Winston.

GARNER, W. R., HAKE, H. W. & ERIKSEN, C. W. (1956). Operationism and the concept of perception. *Psychol. Rev.* **63**, 149–59.

GILINSKY, A. S. (1955). The effect of attitude on the perception of size. *Amer. J. Psychol.* **68**, 173–92.

HAMILTON, V. (1963). Size constancy and cue responsiveness in psychosis. *Brit. J. Psychol.* **54**, 25–39.

HUNT, J. McV. & COFER, C. N. (1944). Psychological deficit. In *Personality and the Behaviour Disorders*, Vol. II (ed. J. McV. Hunt). New York: Ronald.

HUSTON, P. E. & SHAKOV, D. (1948). Learning in schizophrenia. *J. Personality* **17**, 52–74.

JOYNSON, R. B. & NEWSON, L. J. (1962). The perception of shape as a function of inclination. *Brit. J. Psychol.* **53**, 1–15.

LEIBOWITZ, H. W. & PISHKIN, V. (1960). Perceptual constancy in chronic schizophrenia. *J. consult. Psychol.* **24**, 304–14.

LOVINGER, E. (1956). Perceptual contact with reality in schizophrenia. *J. abnorm. soc. Psychol.* **52**, 87–91.

PASCAL, G. R. & SWENSON, C. (1952). Learning in mentally ill patients under conditions of unusual motivation. *J. Personality* **21**, 240–9.

PEREZ, P. (1961). Size constancy in normals and schizophrenics. In *Perceptual Changes in Psychopathology*, ed. W. H. Ittleson and S. B. Kutash, pp. 39–55. New Brunswick, N.J.: Rutgers University.

RAUSCH, H. L. (1952). Perceptual constancy in schizophrenia. *J. Personality* **21**, 176–87.

SCHEFFÉ, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* **40**, 87–104.

STOTSKY, B. A. (1957). Motivation and task complexity as factors in the psychomotor response of schizophrenics. *J. Personality* **25**, 327–43.

TOPPING, G. G. & O'CONNOR, N. (1960). Response of schizophrenics to incentives. *Brit. J. med. Psychol.* **33**, 211–14.

WECKOWICZ, T. E. (1957). Size constancy in schizophrenic patients. *J. ment. Sci.* **103**, 475–86.

(*Manuscript received* 7 *July* 1964; *revised manuscript received* 9 *February* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 335–341
Printed in Great Britain

AN EXPERIMENTALLY INDUCED CONFLICT BETWEEN VISION AND PROPRIOCEPTION

BY RAY OVER
Psychology Department, University of Otago

Settings of a bar so that if felt horizontal were examined under conditions where subjects were given false visual information of the slant of the bar. With a distortion of 15° subjects reported that the bar felt horizontal when it in fact looked horizontal. Although for larger distortions most subjects became aware of a conflict between what was seen and felt, few were able to make accurate settings. The bar tended to be set at a slant between the visual horizontal and the physical horizontal. Settings were variable over trials but there was no trend towards greater accuracy. The data indicate that in making spatial judgements more reliance is placed on visual than proprioceptive input.

The present experiment investigated subjects' capacity to make accurate proprioceptive judgements when presented with false visual information. Proprioception as a sensory system provides precise input relating to the spatial location of a limb. Receptors in the ligaments and joint capsules signal the direction, rate and extent of joint movement, as well as the steady position of a limb (Rose & Mountcastle, 1959). Psychophysical studies indicate fine proprioceptive discriminations in the absence of vision. The horizontal can be accurately judged (Day & Singer, 1964) and slight angular movements of a limb are able to be detected (Piéron, 1952). Vision is similarly a precise locating sense. Most often the two senses provide equivalent information about the location of a limb and the limb feels to be where it looks to be. An intersensory conflict can, however, be induced by falsifying visual input.

In a task of this sort Gibson (1933) required a subject to move his hand along a flat bar which was viewed through a prism that produced visual curvature. The bar was reported to feel curved. Rock & Victor (1964) obtained reports of the shape an object felt to be when the object, grasped by the hand, was viewed through a system that compressed the image along its horizontal axis. Subjects reported that the object felt the shape it looked to be; few subjects were aware of the presence of conflicting sensory information. These data indicate that visual input dominates proprioceptive input when the subject is required to make spatial judgements to a proprioceptive criterion. Neither Gibson (1933) nor Rock & Victor (1964), however, determined whether the dominance by vision is maintained over a wide range of visual distortion values, and they did not consider judgements over a number of trials. The present experiment examined these variables. The task required each subject to make ten settings of a bar so that it felt horizontal while viewing the bar through prisms which make it look to be at a slant other than its actual slant. Distortion values from 0° to 90° were used.

METHOD

Apparatus

The wooden bar which the subject was required to inspect and set was 24 in. long, 3·5 in. deep, and 0·75 in. thick, and was pivoted at the centre. The middle 6·25 in. of the bar was marked by transparent tape which could easily be felt by the subject. In inspecting the bar the subject was required to move the tips of the fingers of his right hand along this centre section. As the arm was not supported, movement of the fingers involved the active movement of the hand and arm. The bar was mounted 1·5 in. in front of a wooden screen 18 in. square. The subject could change the slant of the bar by rotary controls operated through a 6:1 reduction gear by the left hand. A setting of the bar could be read to the nearest 0·25° from an 8 in. protractor fixed at the rear of the screen.

The subject could view the bar only when looking through an eyepiece mounted in a frame 20 in. long and 10 in. high, which was set 14 in. in front of the bar. He could reach the bar by placing his right hand beneath the frame. The eyepiece contained two Dove prisms 0·6 in. × 0·9 in. × 1·88 in., one being fixed in position and the other able to be rotated to any position by the experimenter. With this arrangement a physically horizontal bar could be made to look at any slant from the visual horizontal without a left–right inversion being produced. The mountings in which the prisms were held provided a circular visual field for all degrees of distortion. The limits of the visual field were the centre 6·25 in. of the bar. When the bar was horizontal and there was no visual distortion, the bottom half of the visual field was occupied by the bar, which was black, and the top half by the screen, which served as a white untextured background.

Plan of experiment

Seven conditions of visual distortion were used (0°, 15°, 30°, 45°, 60°, 70° and 90°), termed V_0, V_{15}, $V_{30}, \ldots, V_{90}$ respectively, together with an eighth control condition, P, in which subjects were unable to see the bar. Eight groups, each of eight subjects, were used, i.e. a different group for each condition. Group P made settings without being able to see the bar. The other seven (visual) groups viewed the bar while making settings. For half of each group, except groups V_0 and P, the prisms were arranged so that the left-hand side of the bar, relative to the subject, looked to be sloping upwards when the bar was physically horizontal; for the other half the right-hand side looked to be sloping upwards. Each subject made ten settings.

Subjects

The subjects were high-school students aged between 16 and 18 years. They were not aware of the aims of the experiment.

Procedure

Subjects in all groups were instructed that their task was to set the bar so that it felt physically horizontal.

It was necessary to familiarize subjects with the procedures before testing began. An occluder was used to prevent the subject viewing through the eyepiece. The subject reached beneath the frame. His hand was placed on the bar by the experimenter and moved over the centre span for 15 sec in time with a metronome (60 beats/min). The subject was then allowed to use the rotary controls with his left hand.

Two starting positions (15° on either side of the physical horizontal) were used to obtain test measures; each was used for half of the settings made by each subject. At the beginning of each trial the subject's hand was placed on the bar by the experimenter and the subject was then allowed to look through the eyepiece (except subjects in group K). This was done to ensure that the subject's hand was in contact with the bar whenever he received visual information.

No restriction was placed on the time that subjects could take to make settings. After each setting the occluder was placed into position while the judgement was recorded and the bar set to the starting position for the next trial. On each trial subjects were carefully watched to ensure that they did not close their eyes while making settings. At the end of testing subjects were questioned to determine whether they were aware of a conflict between what was seen and felt, as well as to establish the ways in which they attempted to judge when the bar felt horizontal.

RESULTS

The settings of groups P and V_0 were scored 'U' or 'D', where 'U' was a setting with the right-hand side of the bar, relative to the subject, slanted upwards from the physical horizontal, and 'D' downwards. Settings for the other visual groups were scored '+' or '−', where '+' was a setting from the physical horizontal in the direction of the visual horizontal, and '−' a setting in the opposite direction.

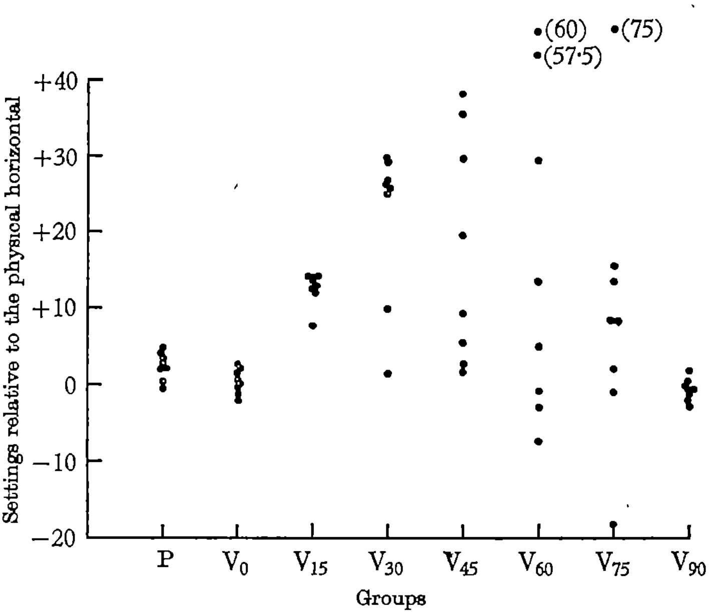


Fig. 1. Settings to the apparent horizontal by each subject on the first trial. Group P proprioception) made settings without vision; an optical system distorted the visual horizontal by 0°, 15°, 30°, 45°, 60°, 75°, and 90° for groups V_0–V_{90} respectively. The values of extremely discrepant settings made by two subjects in group V_{60} and by one subject in group V_{75} are given.

The first comparison between the groups is of settings made on the first trial; the settings of subjects in some groups varied appreciably over trials and the degree of consistency over trials is considered later. Fig. 1 shows the setting made by each subject of each group on the first trial; it indicates that settings have been influenced by the visual distortion conditions. All of the visual groups except V_0 and V_{90} made settings which differed significantly from those made by group P, who made settings in terms of proprioceptive input alone. Inspection indicates that there was a trend towards 'visual' settings (reporting that the bar felt horizontal when it looked horizontal) for visual distortions of up to 30°. For distortions between 30° and 75° 'compromise' settings (where the bar was set at some position between the physical horizontal and the visual horizontal) were most common. Subjects in groups V_0 and V_{90} were as accurate in their settings as subjects in group P. The results thus show that the extent to which proprioceptive judgements were influenced by vision was dependent on the extent of the visual distortion.

The mean variation in settings over the ten trials within subjects in group P provides an index of the degree of inconsistency in judgements which can be attributed to proprioception alone. If the influence of visual information declines over trials and judgements come to be made more in terms of the proprioceptive input, a greater mean variability within subjects over trials would be expected for those groups in which judgements were initially influenced by vision. The standard deviation of his ten settings was calculated for each subject. Table 1 sets out the mean standard deviation for each group and the dispersion of values around this mean. An analysis of variance, applied to a square-root transformation of the data, is also summarized in Table 1. Using Duncan's new multiple range test (Edwards, 1963) all differences other than $P–V_0$, $P–V_{15}$, $P–V_{90}$, $V_0–V_{90}$, $V_{15}–V_{90}$, $V_{30}–V_{75}$, $V_{45}–V_{60}$ were found to be significant ($P < 0{\cdot}01$).

Table 1. *Analysis of within-subject variations in settings for the different groups* ($n = 8$/*group*)

(*a*) Mean variation and S.D. for each group

	Groups							
	P	V_0	V_{15}	V_{30}	V_{45}	V_{60}	V_{75}	V_{90}
Mean	1·60	0·90	2·02	3·61	6·17	7·95	3·79	1·66
S.D.	1·36	0·22	1·53	2·14	4·90	5·46	2·91	0·61

(*b*) Analysis of variance for square-root transformation

Source	Sums of squares	D.F.	*F*	*P*
Between groups	16·92	7	6·37	< 0·01
Within groups	21·11	56	—	—

The results thus show that any variation in settings over the ten trials within subjects in groups V_0, V_{15} and V_{90} was within the range of variation found for group P. All of the subjects in groups V_0 and V_{90} made accurate settings on all ten trials. The subjects in group V_{15} tended to set the bar 15° from the physical horizontal on all ten trials, that is, the visual input was always dominant. Questioning after the experiment indicated that all subjects in this group believed that they had set the bar so that it was physically horizontal; none was aware of any conflict between what was seen and felt.

It has been shown that there is greater variability in settings over trials for subjects in groups V_{30}, V_{60}, V_{75} and V_{90} than can be accounted for in terms of the proprioceptive input alone. It is asked whether this variability represents a trend from 'visual' and 'compromise' settings on earlier trials to accurate settings on later trials. The mean settings made by subjects in each group on trials 1, 3, 5, 7 and 9 are shown in Table 2(*a*); together with the corresponding deviations. After a logarithmic transformation of the raw scores, a test for trend was carried out to determine whether mean settings differed significantly over trials. Table 2(*b*) summarizes the analysis of variance which shows that settings do not differ significantly either over trials or between groups. Hence, there was not an initial reliance on visual input and a later reliance on proprioceptive input. Subjects in these groups made variable settings but did not become increasingly accurate over trials. Only two of the subjects, both in group V_{75}, made accurate settings on all trials and only three, two in group V_{30} and one in V_{75}, made 'visual' settings on all trials. The remaining subjects made 'compromise' settings and they all varied more over the ten trials than any other subject in group P. None showed a consistent trend towards accuracy over the ten trials. Questioning after the experiment indicated that all of these subjects were aware of a conflict between what was seen and felt but that all attempted to set the bar to the physical horizontal.

Table 2. *Analysis of mean settings over trials,* 1, 3, 5, 7 *and* 9 *for groups* V_{30}, V_{45}, V_{60} *and* V_{75}

(*a*) Mean settings and S.D.

		Trials				
Group		1	3	5	7	9
V_{30}	Mean	24·25	21·25	23·75	19·44	21·56
	S.D.	5·81	7·76	7·92	7·95	6·81
V_{45}	Mean	18·37	19·13	18·19	11·47	13·00
	S.D.	14·27	14·09	17·05	13·08	14·98
V_{60}	Mean	19·25	9·31	13·50	13·28	16·91
	S.D.	25·21	10·98	19·91	13·96	20·59
V_{75}	Mean	11·78	12·41	10·34	10·94	10·70
	S.D.	25·50	24·73	25·62	23·45	24·99

(*b*) Analysis of variance for log. transformation

Source	Sums of squares	D.F.	*F*	Level of significance
Visual distortion	1·65	3	1·90	N.S.
Error (*a*)	8·18	28	—	—
Trials	0·03	4	0·00	N.S.
Visual distortion × trials	0·04	12	0·00	N.S.
Error (*b*)	1·93	112	—	—

DISCUSSION

The results show that, although subjects can make accurate settings to the physical horizontal when proprioceptive information alone is available, settings are markedly inaccurate when subjects are presented with veridical proprioceptive and false visual information of the spatial location of the bar. Visual input entirely dominated judgements when the discrepancy was 15°. This finding is consistent with data obtained by Gibson (1933) and by Rock & Victor (1964). For larger discrepancies, however, settings between the physical horizontal and the visual horizontal were generally made. An exception was the group for which there was 90° distortion; all subjects in this group made accurate settings. Subjects who initially made inaccurate settings did not become more accurate over trials.

It is not possible to determine whether the accurate settings made by subjects in group V_0 were in terms of the proprioceptive or the visual input. The results of other groups suggest, however, that these subjects reported that the bar felt horizontal because it looked horizontal. This would imply that the judgements one generally makes of the apparent location of a limb are dependent on the visual appearance of the limb. It is somewhat surprising that greater reliance should be placed on visual input, which has a lower ecological validity than proprioceptive input. Vision, however, is a much less restricted spatial sense than proprioception in that it encodes information about the spatial properties of objects out of reach.

Subjects in group V_{90} made consistently accurate judgements. It is possible that these subjects in fact adopted a visual criterion; that is, reported that the bar felt horizontal when it looked vertical. If this happened it would be expected that the same method would be used by subjects in group V_{75}. Only one subject in this group, however, made a setting of $-15°$ on any trial. Thus in some way subjects were able to make judgements entirely independently of the false information when the distortion was 90° but not when it was 75°.

When questioned after the experiment most subjects reported that they attempted to set the bar at an orientation such that the hand in one direction did not feel to be moving upwards at a slant and in the other direction downwards at a slant. Why is it, however, that this method led to accurate settings in the absence of vision but to errors with visual distortion? Why was vision completely dominant with 15° of distortion but not with larger discrepancies? Why were subjects who became aware that vision was distorted not able to make settings independently of the visual information? There seem to be no ready answers to these questions. Almost all of the inaccurate settings were at some slant between the physical horizontal and the visual horizontal and were 'compromise' settings in this sense; there were very few settings outside this range. It does not seem profitable, however, to use the term 'compromise' except as a label.

The part played by visual and proprioceptive information in judging spatial orientation has been studied by intersensory localization tasks as well as in intersensory conflict situations (Howard & Templeton, 1966). Legge (1965) has found that larger constant errors are made in setting a visible object into alignment with a felt object than in aligning either two visible objects or two felt objects. This finding and the present data both indicate that there are mediating processes involved in translating and correlating information from different modalities. The nature of these processes, and the way they operate to produce error, remains to be established.

The finding that the extent of mislocation induced by visual distortion is not modified over trials seems to conflict with the data from perceptual adaptation studies (Smith & Smith, 1962). It has generally been found that localization responses such as aiming, reaching and making contact are initially controlled by the false visual input but become accurate over trials involving active movement. Adaptation has typically been accounted for in terms of sensory feedback resulting from active motor movement in the presence of the distorted visual information (Held & Freedman, 1963). The present experiment has failed to demonstrate adaptation although active movement of the hand was allowed. The difference in results may be a function of the task required of the subject. In the present experiment the subjects' hand was always in contact with the stimulus and the subject was required to set the stimulus in a given spatial location. In perceptual adaptation studies the subject's task generally is to achieve spatial contact or conjunction with the stimulus. Active movement of the hand changes the apparent spatial location of the hand and parts of the visual field. It is possible that adaptation would have been found in the present case if the subject had been required to achieve contact with the bar before setting it to the apparent horizontal. This would suggest that proprioceptive information alone does not provide the sensory feedback, but the information relating active movement and the degree of spatial conjunction may.

REFERENCES

DAY, R. H. & SINGER, G. (1964). Spatial after-effects within and between kinesthesis and vision. *J. exp. Psychol.* **68**, 337–43.

EDWARDS, A. L. (1963). *Experimental Design in Psychological Research.* New York: Holt, Rinehart and Winston.

GIBSON, J. J. (1933). Adaptation, after-effect and contrast in the perception of curved lines. *J. exp. Psychol.* **16**, 1–31.

HELD, R. & FREEDMAN, S. J. (1963). Plasticity in human sensorimotor control. *Science* **142**, 455–62.

HOWARD, I. P. & TEMPLETON, W. B. (1966). *Human Spatial Orientation.* London: Wiley.

LEGGE, D. (1965). Analysis of visual and proprioceptive components of motor skill by means of a drug. *Brit. J. Psychol.* **56**, 243–54.

PIÉRON, H. (1952). *The Sensations.* New Haven: Yale University Press.

ROCK, I. & VICTOR, J. (1964). Vision and touch: an experimentally created conflict between the two senses. *Science* **143**, 594–6.

ROSE, J. E. & MOUNTCASTLE, V. B. (1959). Touch and kinesthesis. In *Handbook of Physiology, section* 1: *Neurophysiology*, vol. I, ed. J. Field. Washington, D.C.: American Physiological Society.

SMITH, K. U. & SMITH, W. M. (1962). *Perception and Motion.* Philadelphia: Saunders.

(*Manuscript received* 3 *January* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 343–350
Printed in Great Britain

EFFECTS OF SEQUENTIAL ORDERING OF ADDED PAIRS AND OF CORRECTION *V.* NON-CORRECTION PROCEDURES ON PAIRED-ASSOCIATE PERFORMANCE

BY WILLIAM F. BATTIG
University of Maryland

An 8-pair list was learned with fewer errors to criterion under a correction procedure, which equated number of correct responses for each individual pair, than the standard non-correction procedure. Both interpair differences in degree of learning and retention, and variability between subjects, were reduced under the correction procedure. Each of 7 non-correction conditions was then used for a second list of 6 new pairs, learned either alone or in conjunction with the 6 best-learned list 1 pairs (either grouped together, alternated with new pairs, or randomly intermixed), in either constant or varied serial order. Neither sequential grouping nor alternation facilitated list 2 performance, and alternation after non-correction learning of list 1 elicited significantly more errors after the first correct response. Constant serial order for list 2 required fewer trials to criterion than varied serial order.

Several recent experiments have pointed to the extensive involvement of such extraneous cues as serial position (e.g. Battig, Brown & Nelson, 1963) or of various techniques for interpair grouping (Battig, 1966) as means employed by subjects to combat interpair interference during the learning of verbal paired-associate lists (Brown, 1964; Brown & Battig, 1966). One primary purpose of the present study was to assess further the nature and magnitude of effects on paired-associate learning produced by serial-position constancy and sequential arrangement of previously learned and unlearned pairs, and to evaluate any interactions between these variables.

A second focal concern of this experiment was the evaluation of a newly devised paired-associate procedure permitting precise specification and control of level of learning for each individual pair in the list, by requiring that an equal number of correct responses, evenly distributed throughout the course of learning, be made to each pair. This technique, termed the *correction* procedure, owing to its close correspondence with the discrimination-learning procedure of that name, redefines a trial as one correct recall of each pair; each pair that is correctly recalled is removed temporarily from the list until all pairs have been recalled correctly. When used with a recall procedure rather than the standard paired-associate anticipation procedure, this correction procedure effectively mitigates the extensive differences in the level of terminal learning of different pairs that results from the equal number of presentations for each pair under the standard complete-list paired-associate procedure (termed here the *non-correction* procedure). Thus the correction procedure has the potential to eliminate the serious problems associated with the estimation of level of terminal learning in retention and transfer experiments (Underwood, 1964).

The inclusion of the correction/non-correction variable in the present design was not based upon any expected interactions with either serial-position constancy or sequential ordering of paired associates. However, the evaluation of this new technique requires its systematic comparison with the non-correction procedure under manipulated variations of other potentially relevant variables, and the present study is one of a series in which this has been attempted. As the first and most extensive of these correction/non-correction comparisons, the present report also includes several analyses of performance and retention under the correction procedure which contribute towards its fuller understanding and evaluation as compared with the non-correction procedure.

Method

Materials. All subjects learned the same fourteen moderately difficult CVC pairs, selected on the basis of their minimal similarity and equivalent rated learning difficulty (Battig, 1960). The same eight of these pairs constituted the first list for all subjects, whereas the remaining six pairs appeared only in the second list. All pairs were photographed and presented for learning and performance measurement by a Kodak 'Carousel' 2 × 2 in. slide projector.

List 1 *conditions and procedure.* Half of the subjects learned list 1 to a criterion of one errorless trial by the usual non-correction recall procedure: each trial began with the initial presentation of each of the eight pairs at 4 sec rate, followed by a 4 sec blank interval, and then by presentation of each of the eight stimulus terms alone for 4 sec while the subject attempted to pronounce the paired response term.

The correction procedure used for the remaining subjects differed in that a 'trial' was defined as one correct *recall* of each pair rather than its single presentation, thus ensuring that each pair had been recalled correctly the same number of times after each complete correction trial. After the initial presentation and attempted recall of the entire list (which constituted a *complete* non-correction trial), each pair correctly recalled was temporarily removed from the list; the incorrect pairs only were presented again and subsequently tested for recall. Once again, pairs correctly recalled were eliminated. This successive presentation and testing of only the remaining incorrect pairs was continued until the last of them had been recalled correctly. All eight pairs were then restored to the list for the beginning of the next correction trial. Trials were continued until all pairs were recalled correctly following their initial presentation at the beginning of the trial, thus providing a criterion identical to that for the non-correction subjects.

As a check on relative stability of criterion performance under non-correction and correction procedures, correct recall of the entire list on the criterion trial was followed immediately by a second presentation of all eight stimulus terms alone, in a different serial order, for attempted recall of the response terms.

List 2 *conditions and procedure.* Equal numbers of subjects who had learned list 1 under the correction and non-correction procedures were assigned to each of seven conditions for list 2. These conditions differed with respect to the presence and/or the sequential arrangement among the six new pairs of the six old pairs on which subjects had made the fewest errors in learning list 1. The two most difficult pairs for each subject in list 1 were eliminated for use in a subsequent comparison of retention of these pairs after list 2 had been learned by correction or non-correction. Six of the list 2 conditions represented the possible combinations of constant (*Con*) or varied (*Var*) serial order from trial to trial, with grouped (*Gr*), alternation (*Alt*), isolation (*Iso*) conditions of presentation of the new and old pairs in list 2. Under *Gr* conditions, the old pairs always appeared in the first six serial positions for half the subjects, and the last six positions for the remaining subjects. The *Alt* condition provided that each old pair was always adjacent to two new pairs and vice versa, with the old pairs appearing in odd and even numbered positions for half the subjects in each group. Under *Iso* conditions, the six new pairs were presented alone without any old pairs. The seventh *Random* condition provided for unsystematic intermixing of the six old and new pairs which varied both within and between trials.

Learning of list 2 for all subjects followed the list 1 non-correction procedure to the same criterion of one errorless trial.* For *Iso* groups, this criterion trial was followed by the addition of the six old list 1 pairs corresponding to those used in *Gr* and *Alt* second lists, and trials were continued again to a one-errorless-trial criterion. Equal numbers of *Iso* subjects within each group were given post-criterion sequential orders corresponding to the *Gr*-first, *Gr*-last, *Alt*-odd, and *Alt*-even conditions.

---

* Owing to time limitations, two subjects (both in the NC-Var-Alt group) had to be terminated before reaching criterion on the second list. Since both subjects had made only one error on their final trial, and this pair had been recalled correctly on at least one preceding trial, it was assumed for purposes of analysis that these subjects would have achieved criterion on the next trial.

As a further control for *Gr, Alt* and *Iso* subjects under *Con* conditions, six different sequential arrangments of the six new pairs were used equally often for each group, which provided that each pair appeared once in each of the six positions of the sequence, and was preceded and followed once by each other pair. Under *Var* conditions, the particular old and new pairs appearing in each designated serial position were unsystematically varied each time through the list.

Written recall. Immediately after the final list 2 criterion trial, each subject was asked to write down on a prepared written recall sheet as many of the stimulus and response terms for the fourteen pairs as he could remember.

Subjects. The subjects were 168 students at the University of California, Berkeley, 102 of whom were paid for their services while the remainder participated to fulfil an undergraduate psychology course requirement. Three additional subjects, each from a different group, were eliminated and replaced by the next subject to appear in the laboratory, owing to refusal to follow instructions or to procedural or apparatus errors. Each of the two experimenters ran eleven and one of the subjects respectively in each of the fourteen major groups: the correction and non-correction conditions for list 1 combined with the seven list 2 conditions (*Con–Gr, Con–Alt, Con–Iso, Var–Gr, Var–Alt, Var–Iso,* and *Random*). This experimenter variable will not be discussed further, since neither the overall experimenter difference nor its interaction with groups approached significance when tested against the within-groups error term for subjects run by the first experimenter ($F < 1$ in all cases).

Except for the two *Random* groups, which were not run until after all subjects in the other twelve groups had been completed, assignment of subjects to groups was based on a predetermined unsystematic order of assignment which provided that an equal number of subjects had served in each of the twelve groups before any additional subject was assigned to any of these groups. A similar basis was subsequently used for assigning subjects to the correction and non-correction random groups, who were obtained in the same manner and from the same population as the other subjects.

RESULTS AND DISCUSSION

Since the groups representing the seven list 2 conditions were closely comparable in list 1 performance ($F < 1$), and the detailed list 1 analyses involving correction and non-correction procedures were largely independent of the other purposes of the experiment, the list 2 results are presented and discussed first, followed by a separate section devoted to the various correction/non-correction comparisons.

List 2 performance

Results. Comparisons of list 2 performance were based primarily on the six new list 2 pairs, since sixty-seven of the 168 subjects performed perfectly on the six added list 1 pairs, and only 1·4 mean list 2 errors were made on these pairs. Likewise, twenty-one of the forty-eight *Iso* subjects performed perfectly when the six list 1 pairs were added after reaching criterion on the six new list 2 pairs, and very few mean errors were made either on list 1 (0·92) or list 2 pairs (0·88) within the combined list. Thus all list 1 and *Iso* post-criterion list 2 errors were excluded from the list 2 results and analyses to be reported.

Measures of list 2 total errors (failures to recall correctly), trials to criterion, and 'before' errors (preceding the first correct response to each pair) were subjected to complete analyses of variance, as well as two analyses of covariance based upon (*a*) list 1 errors, and (*b*) a multiple covariate including also list 1 correct responses. An analysis of variance was also performed on 'after' errors (errors made after the first correct response to each pair). Neither the total- nor before-errors analyses revealed any difference between groups that approached significance, although their directions were generally consistent with the other measures described below.

Fewer trials were required under *Con* (4·06) than *Var* (4·52) conditions, a difference significant ($P < 0{\cdot}05$) by both analyses of covariance but not by analysis of variance ($P < 0{\cdot}10$). However, neither *Con* nor *Var* trials differed significantly from *Random* groups (4·42). The *Con–Var* differences fell short of significance both for after errors ($P < 0{\cdot}20$) and errors on list 1 pairs ($P < 0{\cdot}10$), and non-significantly fewer before errors were made under *Var* than *Con* conditions ($F < 1$).

Only the after-errors measure revealed any effect of sequential grouping, shown by Table 1 to reflect the inferiority of *Alt* to *Gr* and *Iso* conditions ($F = 3{\cdot}31$; D.F. 2, 154; $P < 0{\cdot}05$). The limitation of this difference to the non-correction procedure led to a highly significant *Gr–Alt–Iso* × C–NC interaction ($F = 6{\cdot}92$; D.F. 2, 154; $P < 0{\cdot}005$). All other list 2 main effects and interactions, as well as the *Gr*-first-last and *Alt*-odd-even comparisons, fell far short of significance for all performance measures.

Table 1. *Mean second-list after errors by Gr–Alt–Iso and C–NC groups*

	Gr	Alt	Iso	Total
C	1·29	0·58	0·58	0·82
NC	0·58	2·67	0·75	1·33
Total	0·94	1·63	0·67	

Discussion. These results are consistent both with the *Con* over *Var* superiority shown by Battig *et al.* (1963) to be limited to later stages of learning, and with the evidence for facilitation of PA learning through the addition of a few previously well-learned pairs to the list (Brown & Battig, 1966). Although the present addition of six such pairs under the *Gr*, *Alt*, and *Random* conditions did not produce superiority to *Iso* conditions, as found by Brown & Battig (1966) for two or four such added pairs, neither did this doubling of number of pairs in the list significantly degrade performance on the six new pairs. When compared with the usual negative effects produced by increasing list length (e.g. Carroll & Burke, 1965), the present equivalence on list 2 pairs for *Iso* as compared with the various added-pair conditions provides impressive evidence of a facilitating mechanism associated with the inclusion of previously well-learned pairs in a new list.

The present inferiority of *Alt* to *Gr* and *Iso* conditions in after errors would appear to argue against a rehearsal interpretation of the facilitation produced by added well-learned pairs, since the systematic alternation of previously learned list 1 pairs with new list 2 pairs should maximize the opportunity to rehearse preceding list 2 pairs during list 1 presentations. Instead, the limitation of this *Alt* inferiority to after errors following non-correction list 1 learning suggests that the presence of well learned pairs increases the use of degree of learning as a basis for interpair grouping, as suggested by Brown & Battig (1966). Because the present correction procedure had produced comparable levels of learning across all added list 1 pairs, this may have enhanced subjects' discrimination between the cues provided by the sequential alternation of old and new pairs under *Alt* conditions, and their ability to use them, to combat interpair interference during the later stages of list 2 learning (Battig, 1966). By comparison, the substantial differences between list 1 pairs in terminal learning level under the non-correction procedure may have obscured and prevented effective use by subjects of the list 2 alternation pattern. A further contributor to such increased differentiation between list 1 and 2 pairs by the correction groups may have been the change for these subjects from the correction to the non-correction procedure for list 2.

Comparisons of correction and non-correction methods

Significantly fewer total list 1 errors were made by correction (28·3) than by non-correction groups (35·3) ($F = 6{\cdot}22$; D.F. 1, 154; $P < 0{\cdot}02$), owing primarily to a marked reduction in errors before the first correct response for correction (20·1) as compared with non-correction conditions (28·3) ($F = 17{\cdot}80$; D.F. 1, 154; $P < 0{\cdot}001$). Although more, but not significantly more correct responses were also made by non-correction (37·9) than by correction groups (36·1) ($F < 1$), the overall superiority of correction in total item presentations fell just short of significance ($F = 3{\cdot}68$; D.F. 1, 154; $P < 0{\cdot}10$). Fewer mean errors were also made by correction (0·30) than by non-correction groups (0·37) on the additional post-criterion recall test, but the difference was not significant ($F < 1$). Only in after errors was non-correction (7·1) not numerically inferior to correction performance (8·2) and this difference was also non-significant ($F < 1$). Finally, the correction procedure led to a substantial reduction in between-subjects error variance which was significant ($P < 0{\cdot}002$) for before errors (C = 78·3; NC = 229·1), and approached significance ($P < 0{\cdot}10$) for total errors (C = 248·3; NC = 395·9).

List 2 performance. Despite the change to a list 2 non-correction procedure for correction groups, list 2 performance was also non-significantly better for correction than non-correction groups for trials, total errors, after-errors, and list 1 errors, whereas correction and non-correction groups were virtually identical in list 2 before-errors ($F < 1$ in all cases). However, except for the previously described interaction with *Gr–Alt–Iso* conditions, no significant list 2 effects resulted from the correction/non-correction variable.

Written recall. Although the written recall results were dominated by a marked overall superiority of recall for items constituting the six list 1 pairs included in list 2 (74·7 %) as compared with the two hardest list 1 pairs not appearing in list 2 (35·4 %), the differences between recall of items from easy (75·2 %) and hard pairs (31·3 %) for non-correction groups were less than the easy (74·2 %) versus hard (39·5 %) difference for correction groups. Neither the mean difference nor a difference between ratios of percentage errors on hard and easy pairs, either for individual item or total pair recall, achieved significance (all P's $\geqslant 0{\cdot}10$). However, significantly more correction (40) than non-correction subjects (24) recalled at least one of the two eliminated list 1 pairs correctly ($\chi^2 = 6{\cdot}46$; D.F. 1; $0{\cdot}025 > P > 0{\cdot}01$). Thus while the differences are not impressive, they are nonetheless consistent with the expectation that list 1 learning of individual pairs was more closely equated under the correction than the non-correction procedure.

Other correction comparisons. Two further analyses were carried out on list 1 performance measures for individual pairs under the correction procedure. The first of these, which is relatively free from bias due to differential selection of subjects, compared the number of errors on trial 2 as a function of the rank order of the eight items in terms of number of errors on trial 1. This relationship, as shown in Fig. 1, clearly indicates a positive correlation between errors for a given pair on the first and second correction trials, although the effect is almost entirely limited to ranks 2–5. A trend analysis of this curve (combining the ranks into four pairs to increase the sensitivity and stability of this measure) revealed a highly significant linear component ($F = 30\cdot65$; D.F. 1, 83; $P < 0\cdot001$), but no differences between ranks 5–8 ($F < 1$) nor between ranks 1–4 ($F = 2\cdot75$; D.F. 1, 83; $P \simeq 0\cdot10$). The superiority of rank 1 to rank 2 was also highly significant ($F = 35\cdot77$; D.F. 1, 83; $P < 0\cdot001$), indicating that if trial 1 was concluded with presentation of a single most difficult pair alone, this produced a substantial improvement in subsequent performance on that pair.

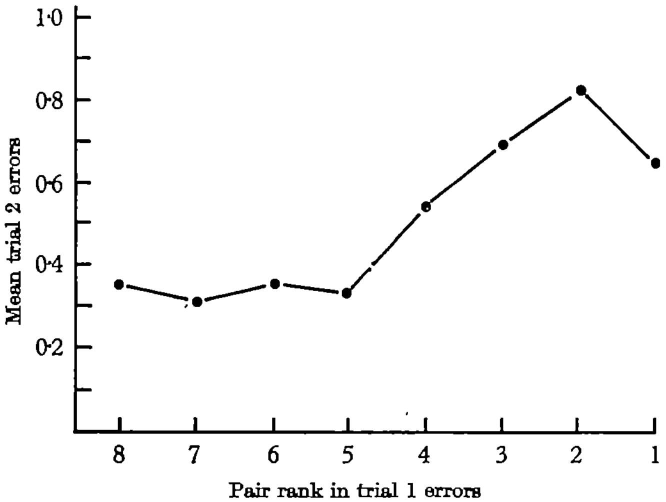


Fig. 1. Mean list 1 errors per pair on trial 2 under C procedure. Pairs are ordered from lowest (8) to highest (1) in terms of number of errors or times presented during trial 1.

Owing to the paucity of errors on trial 2 (4·2) and on all subsequent correction trials combined (4·0), these later trials provided insufficient interpair differences to permit their analysis in the manner depicted in Fig. 1. However, data from all trials were included in a second attempt to compare correction performance as a function of pair difficulty, wherein the number of errors made on each trial was tabulated for each pair as a function of total errors on that pair combined over all preceding trials. This analysis yielded a relation for correction trials 2 and 3 consistent with that depicted in Fig. 1, but the curves for all trials subsequent to trial 3 became flat and virtually indistinguishable. Because of the obvious bias introduced into this analysis by the increasing weight attached to slow-learning subjects, both with increasing trials and increasing numbers of errors preceding each trial, these results are indicative that the level of performance for all pairs has closely approached equivalence on or before the fourth correction trial. Moreover, the direction and magnitude of such differences as were found suggest that these interpair differences might be eliminated simply by requiring one additional correct recall after every three errors made on a given pair. However, the exact nature and generality of application of such a correction for the interpair differences remaining under the correction procedure will necessitate considerable further research.

Discussion. Three conclusions appear warranted by the correction/non-correction comparisons that have been reported: (*a*) the correction procedure produces more rapid and efficient paired-associate learning, and also substantially reduces the magnitude of between-subject differences, thereby minimizing problems created by slow-learning subjects unable to achieve a specified performance criterion; (*b*) both degree of learning and subsequent retention are more closely equated across all individual pairs constituting the list under the correction procedure, although difficult pairs are still learned to somewhat lower levels; (*c*) except for the interaction with conditions of grouping, alternation and isolation (*Gr–Alt–Iso*), both correction and non-correction procedures appear comparable with respect to effects of other manipulated variables. A similar comparability of correction and non-correction procedures has been shown for comparisons between the various standard transfer paradigms after differing levels of list 1 learning (Merikle, Manatis & Battig, 1965), and also in a study evaluating transfer to paired-associate learning from a variety of verbal-discrimination and multiple-choice-recognition pre-training tasks including also variations in paired-associate list length (Battig & Berry, 1966).

However, the significant list 2 C–NC × *Gr–Alt–Iso* interaction, and the present interpretation thereof in terms of the greater equivalence in learning difficulty provided by the correction procedure, should serve to dispel any premature rejection of the correction procedure on the grounds that results obtained with it do not differ from the non-correction procedure. Since few if any variables manipulated in the present and other comparisons of correction and non-correction are demonstrably dependent on degree of learning, the equivalence of their effects appears to indicate merely the comparability of the two paired-associate procedures with regard to general learning processes and phenomena involved, and thus becomes largely irrelevant as a basis for choice between the two procedures. This choice should rest upon whether the unit of experimental manipulation and analysis is the individual pair or the entire list, and the fundamental objection to the non-correction procedure concerns its common inappropriate use where interest is focused primarily upon the analysis of the individual pair. It is here, and only here, that the correction procedure is to be preferred, not only for paired-associate tasks but also for other multiple-response tasks where each item represents a separate unit (e.g. free-recall and verbal discrimination tasks). When inter-item relations are crucial, as in serial learning or studies of context or intralist interference effects (e.g. the present serial-position-constancy and sequential-grouping manipulations on list 2), the correction procedure would be at least as inadequate as is the non-correction procedure when precise control and manipulation of level of learning for each of several items or pairs is critical, as is generally the case in transfer and retention experiments (Underwood, 1964).

This research was supported by Public Health Service Research Grants HD-929 and HD-1062 from the National Institute of Child Health and Human Development. The author is indebted to Jean McMahon for her able assistance in collecting and tabulating the data, and to the Institute of Human Learning, University of California, Berkeley, where the study was designed and the data collected.

REFERENCES

BATTIG, W. F. (1960). Comparison of two methods of scaling nonsense-syllable pairs for ease of learning. *Psychol. Rep.* **6**, 363–6.

BATTIG, W. F. (1966). Evidence for coding processes in 'rote' paired-associate learning. *J. verb. Learn. verb. Behav.* **5**, 177–81.

BATTIG, W. F. & BERRY, J. K. (1966). Effects of number and similarity of pretraining alternatives on paired-associate performance on pretrained and new items under correction and noncorrection procedures. *J. exp. Psychol.* **72**, 722–30.

BATTIG, W. F., BROWN, S. C. & NELSON, D. (1963). Constant *vs.* varied serial order in paired-associate learning. *Psychol. Rep.* **12**, 695–721.

BROWN, S. C. (1964). Interpair interference as a function of level of practice in paired-associate learning. *J. exp. Psychol.* **67**, 316–23.

BROWN, S. C. & BATTIG, W. F. (1966). Second-list paired-associate facilitation produced by addition of previously-learned first-list pairs. *J. verb. Learn. verb. Behav.* **5**, 320–1

CARROLL, J. B. & BURKE, M. L. (1965). Parameters of paired-associate verbal learning: Length of list, meaningfulness, rate of presentation, and ability. *J. exp. Psychol.* **69**, 543–53.

MERIKLE, P. M., MANATIS, J. V. & BATTIG, W. F. (1965). The effects of degree of first-list learning and experimental procedure on transfer under three different paradigms. Paper read at Eastern Psychol. Assn., Atlantic City, N.J.

Underwood, B. J. (1964). Degree of learning and the measurement of forgetting. *J. verb. Learn. verb. Behav.* **3**, 112–89.

(*Manuscript received* 17 *September* 1965)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 351–359
Printed in Great Britain

VANISHING CUES IN PAIRED-ASSOCIATE LEARNING

By WILLIAM M. CHEYNE
Department of Psychology, University of Strathclyde

The typical anticipation method of paired-associate learning was compared with two other methods, where extra cueing was given on initial learning trials and vanished as trials proceeded. The technique of arranging the order of trials, so that successive trials with each pair were spaced progressively farther apart, was shown on two occasions to increase learning. The technique of supplying progressively fewer letters of the response terms as cues gave inconsistent results. Only when cueing was spread over a large number of trials and subjects had prior experience did facilitation occur, and in one case the cues had a detrimental effect on learning. An explanation in terms of 'stimulus selection' is suggested.

In writing programs of the linear type for automated instruction, most investigators accept the principle of controlling the occurrence of errors and correct responses by making use of 'fading', 'vanishing' or 'decreasing cues' during learning. Initially many cues are supplied to the learner to maximize correct responses, but cues are gradually withdrawn as trials proceed until finally the learner is responding in the criterial stimulus situation with no extra stimulus support. Though there is evidence that a vanishing-cues procedure is more efficient than other procedures for teaching a discrimination to pigeons (Terrace, 1963), the basis for the use of this technique in human verbal learning is rational rather than empirical. Various investigators have demonstrated that the vanishing-cues technique can be used with human subjects in a simple matching-to-sample task (Moore & Goldiamond, 1964), in teaching colour names (Taber & Glaser, 1962), or in teaching letter sounds (Popp & Porter, 1960), but in none of these studies was the vanishing-cues technique compared with other procedures where cues were removed suddenly rather than gradually, or where cues were never supplied at all.

Zeaman (1959) points out the similarity between the situation faced by a learner with a linear program and the typical paired-associate learning situation. In both cases the learner is presented with a series of discriminative stimuli to which he makes responses which are immediately confirmed or corrected by the presentation of the correct response. By comparison with the already organized materials which constitute the subject-matter for most programs, however, paired-associate materials have the advantage that variables such as meaningfulness, similarity, and the relation between stimulus and response terms can be more adequately controlled.

The typical anticipation method, where the list of pairs to be learned is shown one pair at a time in different orders, almost maximizes errors during learning and can thus provide a base-line for the study of the effects of cueing techniques. Using no materials other than the particular stimulus and response units to be associated, two vanishing-cues techniques that are commonly used can be studied with paired associates. The first of these, common in programs teaching spelling, foreign vocabulary, or technical terminology, is the method of supplying fewer letters of the response terms as cues, as trials proceed. The second technique is one used frequently but unsystematically in most linear programs, where successive trials with any item are spaced progressively farther apart as learning occurs. The aim of the present investigation therefore was to study the effect of these two vanishing-cues procedures in paired-associate learning, and to compare them with the anticipation procedure providing minimal cueing.

EXPERIMENT I

In this experiment groups having (*a*) cueing from the order of the trials, (*b*) parts of the response as cues, and (*c*) a combination of both these methods, were compared with a group having no extra cueing; recall was tested after subjects had worked unpaced for a fixed period of time.

Method

Materials

One list of fifteen pairs was learned by all subjects. Stimulus terms were names of countries, and response terms were tri-consonant syllables of 50 % association value according to Witmer (1935). The response terms were chosen so that their intra-list similarity was as low as possible, and in particular so that no two syllables had the same letter in the same position.

Apparatus

The apparatus consisted essentially of a box with two paper rolls inside it. By turning a roller, subjects moved the two tapes forward synchronously behind a Perspex window, revealing the materials on the left-hand tape, and writing their responses on the right-hand tape through an aperture in the window. Stimuli and responses moved under clear Perspex, where they could be seen but not altered, before moving out of sight under opaque masking.

Experimental plan and conditions

Four experimental conditions formed a 2 × 2 design contrasting no cues with decreasing cues in combination with randomized or programmed orders of presentation. A difference group of subjects was used for each of the four conditions.

The *no cues* condition (NC) used what was essentially a self-paced anticipation/confirmation procedure. Subjects were shown the stimulus term of each pair alone and required to supply the response if possible; they were then shown the response term for confirmation or correction. The complete list of pairs was repeated in this way as often as required in a different random order each time.

The *decreasing cues* condition (DC) differed from the NC procedure in that letters of the response terms were shown alongside the stimulus terms as cues on initial trials. On the first *two* trials with each pair the complete response term was shown alongside the stimulus term, on the next *two* trials two letters of the response term were supplied as cues, and on the next *two* trials only one letter was given with the stimulus term. From the seventh trial onwards with each pair, the stimulus term was presented with no letters of the response term as cues, and the procedure became identical to that for the NC group. The letters given as cues were counterbalanced over pairs and trials to ensure that first, second and third letters of the response terms were given equally often as cues.

The *programmed order* condition (O) gave no letter cues, but the order of trials was arranged so that pairs were added in as trials proceeded, and the first nine trials with each pair were spaced progressively further apart, i.e. between successive trials with any particular pair, the number of trials with other pairs increased as trials proceeded. When all pairs had been added in this way, the order of the complete list was randomized and the procedure became identical to the NC procedure.

The fourth condition (ODC) had both the *programmed order* of trials as in the O condition, and *decreasing cues* like the DC condition. Subjects were given all three letters of the response term as cues on the first *two* trials with each pair, two-letter cues in the next *three* trials, one-letter cues on the next *three* trials, and no cues on the ninth and all succeeding trials.

Subjects

The eighty subjects were first-year psychology students at the University of Aberdeen. They had had no previous experience of verbal learning experiments. Twenty subjects were assigned to each of the four conditions.

Procedure

On each trial with each pair under all experimental conditions subjects were instructed (i) to move the tapes on to reveal the stimulus term (with or without letters of the response term alongside it); (ii) to write response on answer tape; (iii) to move the stimulus term and response under the Perspex and reveal the response term; (iv) to copy down the response term if response was incorrect; (v) to move the pair of terms and the responses out of sight revealing the next stimulus term.

After instructions and practice with three extra pairs to ensure that they understood the procedure, subjects were given a 15 min learning session with the main list of fifteen pairs. Learning was followed immediately by a paced recall test with no confirmation of responses; subjects were shown each stimulus term for 10 sec and required to write down the appropriate trigrams. In a backward recall test, subjects were then shown each of the trigrams for 10 sec and required to supply the appropriate stimulus terms.

Finally subjects were given a sheet with the list of pairs, and were asked to write down all the associations to the materials which had occurred to them during learning; they were cross-examined by the experimenter to ensure that all associations were listed, and that the reported associations could be classified according to whether they could mediate in response integration, mediate between stimulus term and response term or perform both these functions.

RESULTS

(*a*) *Recall scores*

The mean numbers of correct responses made by each experimental group in the forward and backward recall tests are shown in Table 1. An analysis of variance of the recall scores (type III design, Lindquist, 1953, pp. 281–4) showed three significant sources of variance (cf. Table 2): the interaction between the two methods of cueing (O × DC), the main effect of type of recall test (R), and the interaction between type of recall test and the decreasing cues technique (R × DC).

Table 1. *Mean numbers of pairs correctly recalled in forward and backward recall tests in Expt. I*

	Forward recall		Backward recall	
Group	Mean	S.D.	Mean	S.D.
NC	7·80	4·35	10·00	4·00
DC	9·05	3·80	11·55	2·57
O	10·05	3·63	11·15	2·80
ODC	6·85	4·55	9·35	2·58

Analysis of the interaction between methods of cueing showed that group ODC recalled significantly less than group O ($t = 3{\cdot}16$; D.F. 38; $P < 0{\cdot}01$) but the difference between groups DC and NC is not statistically significant. Group O recalled significantly more than group NC ($t = 2{\cdot}14$; D.F. 38; $P < 0{\cdot}05$), but groups ODC and DC showed a sgnificant difference in the other direction ($t = 2{\cdot}78$; D.F. 38; $P < 0{\cdot}02$). It appears therefore that the programmed order alone resulted in more learning than the method of presenting the complete list in changing orders. The addition of decreasing cues to this programmed order eliminated this superiority, though the decreasing cues had no detrimental effect when added to the random order.

The significant variance between recall tests reflects the overall superiority of backward over forward recall (which came first), though the interaction with the decreasing cues procedure shows that the improvement with backward recall is greater for groups with decreasing cues (DC and ODC) than for the groups with no letter cues (NC and O).

Table 2. *Analysis of variance of effects of type of recall test* (*R*), *random versus programmed order* (*O*), *and no cues versus decreasing cues* (*DC*) *on recall scores in Expt. I*

Source	S.S.	D.F.	M.S.	*F*	*P*
		Between subjects			
O	3·025	1	3·025	—	—
DC	13·225	1	13·225	0·518	—
O × DC	156·025	1	156·025	6·117	< 0·025
Error (between)	1938·500	76	25·506	—	—
		Within subjects			
R	168·100	1	168·100	159·035	< 0·001
R × O	3·603	1	3·603	3·009	—
R × DC	6·400	1	6·400	6·055	< 0·025
R × D × DC	2·497	1	2·497	2·361	—
Error (within)	80·400	76	1·057	—	—
Total	2317·775	159	—	—	—

Table 3. *Mean numbers of trials completed and mean percentages of trials on which subjects failed to anticipate correct responses for each group in Expt. I*

	Trials		Percentage errors	
Group	Mean	S.D.	Mean	S.D.
NC	95·10	26·0	75·3	17·1
DC	88·70	21·1	20·8	11·6
O	110·00	26·7	39·2	21·0
ODC	110·25	25·1	23·9	19·6

(*b*) *Trials*

Since subjects worked unpaced for a fixed length of time, different numbers of trials were completed by subjects both within and between groups. Table 3 shows the mean numbers of trials completed in each group. An analysis of variance of the number of trials completed shows significant variance due to the ordering of trials ($F = 10{\cdot}61$; D.F. 1, 76 ; $P < 0{\cdot}005$)—the programmed orders result in faster working —but no other significant sources of variance.

(*c*) *Errors*

The effect of the cueing procedures on the number of errors which occurred is of interest because cueing is usually considered to be a technique which reduces errors. Since subjects completed different numbers of trials, the only measures by which groups could be compared were the percentages of errors made by subjects, and the mean percentages for each group are shown in Table 3. The variance within groups proved to be homogenous (Bartlett's test gave $\chi^2 = 6{\cdot}56$; D.F. 3; $P > 0{\cdot}10$), and an analysis of variance of the error rates was carried out. Both the main effects and the interaction between cueing methods were highly significant (the F ratios being 24·7, 77·8, 17·4; D.F. 1, 76 and $P < 0{\cdot}001$ in all three cases).

The addition of letter cues in the first six trials reduced errors from 75 % in Group NC to 21 % in group DC ($t = 8{\cdot}15$; D.F. 38; $P < 0{\cdot}001$). Comparison of groups NC and O showed that the programmed order almost halved the error rate ($t = 4{\cdot}11$; D.F. 38; $P < 0{\cdot}001$), but group ODC did not make significantly fewer errors than either group O or group DC.

If the number of occasions when subjects made any attempt to anticipate the response term is noted (whether the anticipation was a correct response or an overt error), there are further clear differences between the groups. Group NC made overt responses on 66·8 % of trials compared with 91·1 % in group O. The groups with decreasing cues almost always made overt responses, with 98·9 % and 99·6 % overt responses in groups ODC and DC respectively. In group DC, nine out of the eleven omissions which occurred in 1774 trials were made by a single subject. It is clear that the cueing procedures, particularly decreasingly cues, result not only in a tendency for subjects to make more correct responses, but in a general tendency for subjects to make anticipations rather than omissions.

(*d*) *Mediating responses*

On the basis of the subjects' reports, the pairs were classified into four categories according to the types of mediating associations that occurred: M-types are pairs where a mediating response occurred which could completely integrate the response term and link it with the stimulus term; S-types are pairs where the mediating response could link one or two letters of the response term to the stimulus term; R-types are pairs where the mediating response could assist only in response integration, and X-types are the remaining pairs for which no mediating associations occurred. The mean numbers of M-, S-, R- and X-types occurring in each group are given in Table 4.

Table 4. *Mean numbers of pairs categorized according to types of mediating responses which occurred during learning in each group in Expt. I: see text for definitions of types M, S, R, and X*

	Type			
Group	M	S	R	X
NC	3·10	0·95	4·60	6·35
DC	4·45	1·85	3·85	4·85
O	4·60	2·10	2·20	4·10
ODC	1·35	1·40	3·35	7·35

Independent analyses of variance of the number of M-types, (S+R)-types and X-types in the various groups show a significant interaction between methods of cueing in the occurrence of both M-types ($F = 13{\cdot}92$; D.F. 1, 76; $P < 0{\cdot}001$) and X-types ($F = 8{\cdot}46$; D.F. 1, 76; $P < 0{\cdot}005$), but no other significant sources of variance. The interactions reflect the differences between groups ODC and O in numbers of M-types ($t = 3{\cdot}49$; D.F. 38; $P < 0{\cdot}01$) and X-types ($t = 2{\cdot}42$; D.F. 38; $P < 0{\cdot}05$), while the corresponding comparisons of groups DC and NC show no significant differences. Group ODC, with both decreasing cues and the programmed order, reported fewer pairs learned completely with the use of mediating associations and more pairs for which no mediating responses occurred.

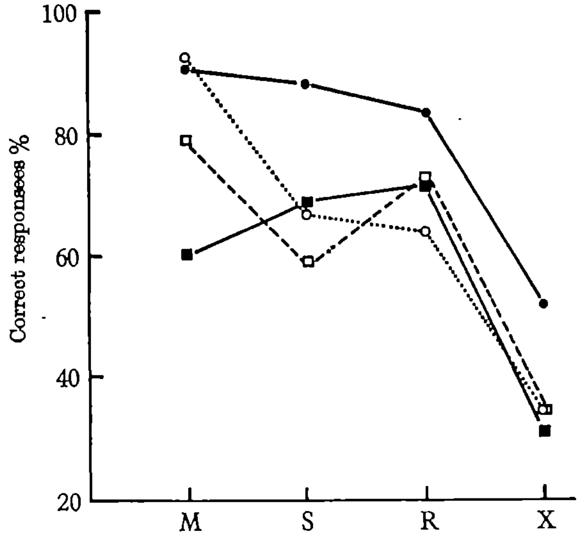


Fig. 1. Percentage recall for pairs with different types of reported mediating responses in Expt. I. ●—●, O; ○---○, ODC; □---□, ■—■, NC.

(*e*) *Mediating responses and recall*

The relation between recall and reported association was studied in view of the possibility that this might be affected by the cueing procedures, e.g. more efficient learning with cues could reduce the difference between pairs learned with and pairs learned without associative aids. This relation is seen most clearly when the percentage of pairs of each type recalled in the forward recall test is calculated for all subjects within each group combined (cf. Fig. 1). Apart from the expected finding that pairs for which mediating associations occurred are recalled more frequently in all groups, there is a clear interaction. The previously found significant difference in recall between groups O and NC occurs for all pairs whether or not mediation occurred. The similar significant difference between the O and ODC groups does not occur, however, with M-type pairs where the mediating associations can both integrate the response term and link it with its stimulus term. Thus the detrimental effect of adding cues to the programmed order occurred only for pairs which were learned wholly or partly without the use of associative aids.

EXPERIMENT II

In this experiment the O, DC and NC procedures were repeated using some of the same subjects with two procedural modifications: all subjects completed the same number of trials, and cueing was revised to reduce errors further.

Method

Materials

A second list of fifteen countries with 50 % AV tri-consonant syllables was presented in the same apparatus. The three groups of eight subjects were twenty-four subjects who had taken part in Experiment I thirteen weeks previously and had just completed a retention test for that experiment.

Experimental conditions

The NC condition was identical to the previous NC condition except that all subjects were given 10 trials with each pair.

The DC condition also included 10 trials with each pair, but letters of the response term were supplied alongside the stimulus term as cues so that three-letter cues were given on the first *three* trials, two-letter cues were given on the next *three* trials, one-letter cues on the next *three* trials, and no cues were given on the last trial with each pair.

The O condition: subjects were given 150 trials with the order of trials arranged so that successive trials with each pair were spaced more closely together than in the previous O condition. Since pairs were added in as trials proceeded, subjects had different numbers of trials with each pair, varying from eleven trials with initial pairs to only seven with the last pair added in.

Subjects were given practice, a learning session for which the time to complete the 150 trials was recorded, and a paced recall test allowing 6 sec in which to supply each trigram. Mediating associations were recorded and categorized as in Expt. I.

Results

The mean scores in the recall test were 9·6, 13·4 and 13·1 for groups NC, DC and O respectively. A Kruskal–Wallis non-parametric analysis of variance by ranks showed significant differences between groups ($H = 6{\cdot}08$, $P < 0{\cdot}05$); and both groups O and DC recalled significantly more than group NC ($P < 0{\cdot}02$ for Mann–Whitney U tests). There were no significant differences in the times taken to complete the 150 trials, nor in the number of reported mediating responses in the three groups. The error rates in the groups DC and O were reduced with the increased cueing to 7 % and 9 % respectively, whereas the error rate remained at 53 % in group NC. The corresponding percentages of anticipations (correct responses + overt errors) were 99·1, 99·0 and 79·0 respectively.

DISCUSSION

The programmed order, as expected, increased correct responses during learning and increased immediate recall. Since this result occurred in both Expts I and II, it is not dependent on the number of trials completed, or on the speed of working, and the difference is not related to the number of mediating associations which occurred.

The decreasing cues technique, however, gave inconsistent results over the three occasions on which it was tested. Only in Expt II, where cueing was spread over nine of the ten trials and where subjects had previous experience, did the addition of decreasing cues result in the expected increase in recall. In the case of the ODC group in Expt I, the letters supplied not only had a detrimental effect on recall, but they did not significantly reduce errors, i.e. the parts of the response did not in fact act as cues. Any explanation of this result must necessarily be *ex post facto*, but it is certainly clear that the recall scores are not related directly to the number of correct responses or errors made during learning.

The phenomenon of 'stimulus selection' which occurs with context cueing (Underwood, 1963) suggests a hypothesis that could explain these results. When the stimulus terms in paired-associate learning are trigrams with coloured borders, stimulus selection may occur so that the functional stimuli for subjects' responses during learning are the colours alone, with the result that very little association is formed between the trigrams and the response terms (Underwood, Ham & Ekstrand, 1962; Jenkins & Bailey, 1964). If such selection occurred when the cues were parts of the response, so that the functional stimuli for subjects' responses during learning were the letters of the response term and not the countries, the decreasing cues procedure could have a detrimental effect on learning since subjects would not respond adequately to the critical stimulus terms (the countries) until the cues were eliminated. This effect would be greater for the ODC group in Expt I which had cueing spread over more trials than the DC group in the same experiment. The smaller number of reported associations mediating between the stimulus term and the response term in the ODC group supports this hypothesis, i.e. that subjects in this group were attending less to the critical stimulus term during learning. Since the DC group in Expt II learned more than the corresponding NC group, after having letter cues for nine out of ten trials with each pair, there is no evidence in this case for stimulus selection. It is to be noted, however, that not only had the subjects in Expt II had previous experience with and without letter cues during learning in Expt I, but immediately prior to taking part in Expt II these subjects completed a relearning test for the previous experiment, involving extended practice with no cues, i.e. with the countries alone as stimulus terms. Both these factors would be expected to influence subjects to respond adequately to the countries in Expt II and thus eliminate any selectivity that could have occurred otherwise. Further experimentation with naïve subjects is needed to settle this question. It is relevant also that the procedure which resulted consistently in increased recall, i.e. the programmed order condition O, is one in which it is almost impossible for subjects to respond to the 'cues' (the ordering of trials) and not to the critical stimuli.

Whatever the explanation for the decreasing cues results, it is clear that the general principle, that error reduction during learning will increase learning as measured by immediate recall, is not necessarily valid, even when cueing is gradually reduced throughout learning.

The superiority of backward over forward recall in Expt I is of interest mainly because this result goes against the general trend of findings from previous tests of stimulus recall (Morikawa, 1959). The result appears to be due to the nature of the materials and to the fact that partially correct responses were counted as errors in the forward recall test. Subjects, particularly when given parts of the response term as cues, might be expected to associate one or two letters of some response terms with the appropriate countries before complete response integration has occurred. In the forward recall test such an association would be given no credit, but owing to the minimal intra-list similarity among response terms, this partial learning might be sufficient to enable recall of the highly integrated stimulus terms, and thus inflate the backward-recall scores. The greater superiority in backward recall in the groups with decreasing cues provides further evidence that this is the case.

REFERENCES

JENKINS, J. J. & BAILEY, V. B. (1964). Cue selection and mediated transfer in paired-associate learning. *J. exp. Psychol.* **67**, 101–2.

LINDQUIST, E. F. (1953). *Design and Analysis of Experiments in Psychology and Education.* Boston: Houghton Mifflin.

MOORE, R. & GOLDIAMOND, L. (1964). Errorless establishment of visual discrimination using fading procedures. *J. exp. Anal. Behav.* **7**, 269–72.

MORIKAWA, Y. (1959). Functions of stimulus and response in paired-associate verbal learning. *Psychologia*, **2**, 41–56.

POPP, H. & PORTER, D. (1960). Programming verbal skills for primary grades. *Audio-vis. Commun. Rev.* **8**, 165–75.

TABER, J. I. & GLASER, R. (1962). An exploratory evaluation of a discriminative transfer learning program using literal prompts. *J. educ. Res.* **55**, 508–12.

TERRACE, H. S. (1963). Errorless transfer of a discrimination across two continua. *J. exp. Anal. Behav.* **6**, 223–32.

UNDERWOOD, B. J. (1963). Stimulus selection in verbal learning. In *Verbal Behaviour and Learning: Problems and Processes*, ed. C. N. Cofer and B. S. Musgrave, pp. 33–48. New York: McGraw-Hill.

UNDERWOOD, B. J., HAM, M. & EKSTRAND, B. (1962). Cue selection in paired-associate learning. *J. exp. Psychol.* **64**, 405–9.

WITMER, L. R. (1935). The association value of three-place consonant syllables. *J. genet. Psychol.* **47**, 337–60.

ZEAMAN, D. (1959). Skinner's theory of teaching machines. In *Automated Teaching: the State of the Art*, ed. E. H. Galanter, pp. 167–76. New York: Wiley.

(*Manuscript received* 3 *February* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 361–373
Printed in Great Britain

TRANSFER IN CATEGORY LEARNING OF YOUNG CHILDREN: ITS RELATION TO TASK COMPLEXITY AND OVERLEARNING

By ANN M. CLARKE and G. M. COOPER
Department of Psychology, The University of Hull

An investigation with normal pre-school children is reported which confirms and extends earlier findings on subnormal subjects. That transfer is related to the complexity of intervening training is again demonstrated across widely different tasks. In addition, evidence is offered for the interaction of overlearning with complexity, and the possible relevance of transfer in cognitive development is briefly discussed.

A series of experiments by Clarke & Cooper (1966) suggested that transfer in imbeciles on visual sorting tasks was related to the complexity of the intervening training. To test this theory further, Clarke, Cooper & Henney (1966) carried out an experiment on imbecile children, aged 12, in which three conditions of training of varying difficulty were used and where the training and transfer tasks differed widely in content. The hypothesis was strongly confirmed; not only did those groups trained on the two more difficult tasks show more transfer than those trained on an easy task or than the untrained control subjects, but within the groups the initial low scorers (who profited more by the intervening training, and for whom the transfer task was more difficult) showed significantly greater gains than initial high scorers.

In the present paper new experiments, based on the previous experiments but using normal pre-school children, are reported. Section I describes a repetition of the earlier research (slightly modified in the light of experience) to test the hypothesis that a relation between task complexity and transfer would also hold with biologically normal, naïve subjects and across widely different tasks. In section II an experiment is described which tests the ability of young children to apply categories acquired in one situation to new exemplars of the same categories. Section III throws light upon a finding in two experiments that, with complex training tasks of different levels of difficulty, transfer did not differ in amount, although it was considerable. Finally, some theoretical problems are considered.

I. TRANSFER BETWEEN DISSIMILAR TASKS: ITS RELATION TO TASK COMPLEXITY: EXPERIMENT I

The first experiment attempted to replicate with normal children the findings of Clarke *et al.* (1966) on imbeciles.

Method

Experimental design

This was identical with that of Clarke *et al.* (1966) except for an addition at the end. Four groups of subjects were equated on the basis of pretest performance on the transfer task, *T*, and on other variables. Thereafter, the four groups were randomly allocated to three training conditions, I, II, III and control, for twenty-four spaced trials. After retest on *T*, nine direct learning trials on this material were given, making a total of eleven trials on T for all four groups. In addition, the subjects were tested at the end of the experiment on two further tasks in which the conceptual basis for sorting was the same as in T, but with different exemplars.

The tasks

All tasks involved sorting response cards, 3 in. × 3 in., into five categories represented by stimulus cards.

As before, the transfer task, T, consisted of black-and-white photographs of objects belonging to five common but rather wide conceptual categories: human beings, animals, furniture, tableware and clothes. A different set of photographs was used, however, in an attempt to make the task slightly more difficult than formerly. The stimulus cards were: baby, pig, cup, sofa, dress. The twenty response cards in order of presentation were: sheep, man, cupboard, knife, boy, bed, plate, shirt, girl, fork, cat, shoes, chair, spoon, hat, horse, table, dog, woman, gloves.

The three training tasks again involved sorting geometric shapes, but entirely new material was used in order to make the two difficult tasks more difficult and the easy task even easier. The new packs were devised in the light of knowledge gained from the analysis of the types of error made by the imbeciles in the previous experiment. In order to guard against possible fatigue and monotony, only fifteen rather than twenty response cards were used during training on these tasks.

Table 1. *Data relating to variables on which groups were matched*

	Group I	Group II	Group III	Control	F
n	8	9	9	13	
Mean age in months	51·63	52·55	54·44	52·77	< 1
Mean pretest score, task T	6·75	6·89	7·11	7·17	< 1
Mean pretest errors, task T	30·62	31·00	29·33	30·23	< 1
No. of boys	5	5	6	8	

Training task I (easy) involved straight matching of five large outline shapes, as distinct from each other in form as possible. They were: cross, star, circle, equilateral triangle and crescent.

Training task II (fairly difficult) involved sorting outline shapes all of which had one curved side. The stimulus shapes were uniformly small, and the response shapes varied from them and from each other in size, orientation and position on the card.

Training task III (difficult) involved sorting outline shapes of different quadrilaterals: squares, rectangles, parallelograms, trapezia and irregular shapes. The figures also varied in size, orientation and position on the card.

Control conditions involved attendance of subjects on alternate days in the experimental room for conversation and social contact with the experimenters.

Subjects

The subjects were children drawn from two local nursery schools, selected on the basis of age, score and number of errors on first performance of task T and their willingness to co-operate. In order to achieve homogeneity of groups, the children were all in their fifth year, except for one child aged 3 years 10 months, and all had an initial score below 13/20 on task T. In order to avoid any possible bias from socio-economic background or method of instruction in the two schools, care was taken to include a similar proportion of children from each school in each of the four groups. Originally each group contained nine children, who had been randomly assigned to the different experimental conditions. After the start of the experiment, however, four more children meeting the criteria for inclusion in the experiment became available and were added to the control group. One child who had been assigned to group I left the area late in the experiment, and could not be replaced. Table 1 gives a summary of the initial data on the groups. No significant differences were found among the initial means (all F ratios being less than one), and the variances were homogeneous.

Procedure

The apparatus consisted of a panel holding the five stimulus cards, each with a box below, the floor of which could be lowered by levers at the side. A red light bulb was fixed in each box immediately beneath the stimulus card.

Each subject was put at ease on entering the experimental room. He was seated in front of the apparatus and verbal instructions were kept to a minimum. Initially two cards, not subsequently used in the experiment, were slowly placed in their correct compartment, the red light flashed and the card dropped out of sight. The subject was then told 'I want you to put all the pictures I give you into their proper places'.

Whenever a correct response was made the red light came on, the experimenter said 'Good' and the card disappeared. If the response was incorrect, the experimenter said 'No', the red light failed to come on and the trap-door remained closed. The card was returned to the subject who was asked to try again. In the rare cases when a correct response was not obtained after five attempts, the experimenter placed the card in the correct compartment and the next response card was offered.

Note was taken of each response and the precise location of each error. In order to minimize the possibility of rote learning, the positions of the stimuli on the panel were randomized on every trial. In the case of the geometric material the order of presentation of the response cards was reversed each time.

Results

The results are given in terms of first scores (i.e. number of placements correct first time) throughout. Errors followed the same general pattern, but do not represent such a stringent measure of task efficiency.

During training on the geometric tasks, group I (easy training) achieved perfect performance on trial 1 and then proceeded to twenty-three further trials of practice (or 'over-learning') on this task; members of group II (fairly difficult training) started with a mean score of 7·00, reached an asymptote at trial 6 (score 14·11/15), and thereafter were overlearning the task for the remaining eighteen trials. The learning of group III (difficult training) in contrast, proceeded much more gradually, starting with a mean score of 4·11/15 and ending at trial 24 with a mean score of 11·56/15.

Two measures of transfer were taken to establish the effect of training with the geometric tasks on task *T*: (1) the differences in scores from pretest to immediate retest after the intervening training; (2) the relative level of acquisition of task *T* over the ten trials of learning. The data were treated statistically by analyses of variance and Tukey's method of comparison of means (Ryan, 1959). Since the groups were not of equal size, the smallest number was used to compute the standard error.

Table 2 shows the immediate pretest-retest score differences for the groups on *T*. The pretest/retest gains made by the children who had undergone the more difficult training (groups II and III) were on the average over twice as great as those made by the group who had experienced easy training (group I) and by the control group. The former two groups did not differ significantly from each other, nor did the latter two. Fig. 1 shows graphically the positions of the four groups at pretest, retest and thereafter during the nine additional training trials on *T*.

Table 3 gives the results of an analysis of variance on the scores over ten trials (retest plus nine) on task *T* following training on the different geometric tasks. In spite of homogeneity of the variances, each of the three experimental groups contained one subject whose scores lay consistently well below the mean for the group

Table 2. *Comparison of groups on pretest/retest difference scores on* T

Analysis of variance

Source	S.S.	D.F.	M.S.	*F*	*P*
Between groups	79·19	3	26·40	4·23	< 0·025
Residual	218·25	35	6·24	—	—

Tukey comparison of means

Group	II	III	I	Control
Means	+ 5·11	+ 4·78	+ 2·12	+ 2·08
II	—	N.S.	< 0·05	< 0·05
III	—	—	< 0·05	< 0·06
I	—	—	—	N.S

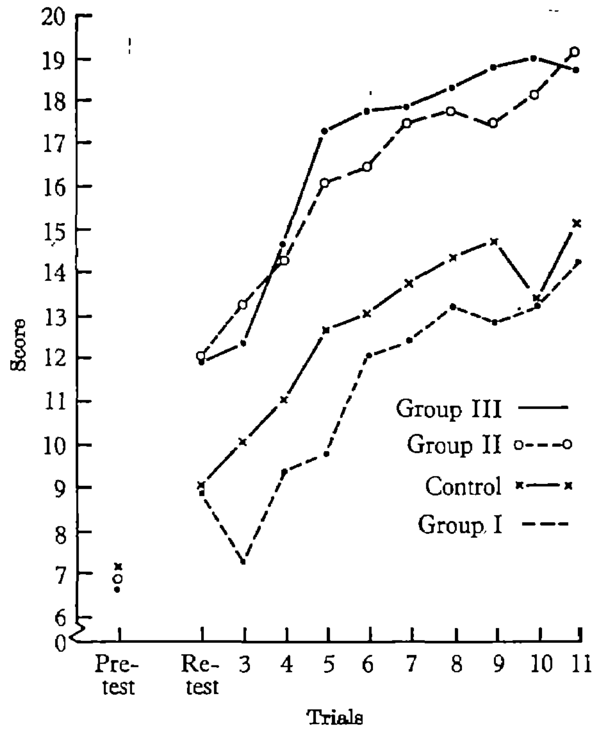


Fig. 1. These curves show the average scores per group on the sorting of twenty photographs into five common conceptual categories. Four groups, equated on pretest score were then allocated to intervening training on geometrical sorting tasks of differing complexity or to control conditions. All groups were then retested on the original material and given nine further acquisition trials. The graph shows the effects of the differing complexity of the intervening training.

(i.e. poor learners), while the control group contained one child with a pretest score of 2 who appeared to show sudden 'insight' on trial 3, her score rising from 6 to 16 and thereafter remaining consistently superior to the rest of the group. This slightly weighted the scales against our hypothesis, which was nevertheless, at least in part, strongly confirmed.

Although each of the more difficult task groups (II and III) differed significantly from the control group and from group I trained on the easy task, they did not differ significantly from each other, either on pretest/retest differences, as already mentioned, or on scores gained during the subsequent learning of *T*. Exactly the same finding had occurred in the previous experiments on imbeciles, and this time therefore a separate investigation was carried out to explore one of the possible reasons for the lack of differential transfer from tasks widely different in level of difficulty. This investigation is discussed below in § III.

Table 3. *Comparison of groups on ten trials on task* T

Analysis of variance

Source	S.S.	D.F.	M.S.	*F*	*P*
Between groups	1804·80	3	601·60	36·68	< 0·001
Residual	6330·24	386	16·40	—	—

Tukey comparison of means

Group	III	II	Control	I
Means	16·57	16·11	12·70	11·27
III	—	N.S.	< 0·01	< 0·01
II	—	—	< 0·01	< 0·01
Control	—	—	—	0·1 > *P* > 0·05

Table 4. *Comparison of groups on last four trials only, of ten acquisition trials on* T

Analysis of variance

Source	S.S.	D.F.	M.S.	*F*	*P*
Between groups	743·81	3	247·94	17·85	< 0·001
Residual	2110·81	152	13·89	—	—

Tukey comparison of means

Group	III	II	Control	I
Means	18·56	17·97	14·40	13·28
III	—	N.S.	< 0·01	< 0·01
II	—	—	< 0·01	< 0·01
Control	—	—	—	N.S.

The only unexpected finding was a significant inferiority of group I (easy training task) to the control group during the early stage of learning *T*. This training task had been devised to be so easy that no positive transfer could reasonably be predicted, and it had been anticipated that members of this group would not differ in their performance from control subjects. Two possible explanations are suggested: (*a*) that the members of group I, even though matched with the other groups, were innately inferior learners to the control subjects; or (*b*) that, in the sense of Helson's (1959) theory of adaptation level, twenty-four trials on an easy task had made it harder for these children to adapt to the demands of a difficult task than they had for those who had experienced no specific training of any kind. In view of the initial matching of children into groups and the random allocation of groups to type of training, the first interpretation seemed very unlikely. If the second explanation were correct, it seemed probable that the children would overcome their initial handicap during the course of learning and should be equal to the control subjects towards the end of the learning period. A separate analysis was carried out on the scores for the last four trials of the four groups. The results are presented in Table 4. Over the last four trials of training on T, group I did not differ significantly from the control group, thus supporting the possibility of negative transfer for which there was a fairly rapid compensation. Groups II and III were by this time regularly gaining perfect or virtually perfect scores, the means being slightly reduced by the one low scorer in each group. For the majority of them, at least, task T appeared to have too low a ceiling.

Conclusions

The earlier main finding using imbecile subjects has been confirmed with normal pre-school children; it appears that complex training tasks produce considerable transfer to widely different tasks. The relation between complexity and transfer may not, however, be a simple one, for although groups II and III (fairly difficult and difficult training, respectively) both showed impressive transfer, this was similar in amount. Some hypotheses to account for this finding are discussed in section III.

II. TRANSFER BETWEEN SIMILAR TASKS: ITS RELATION TO OVERLEARNING: EXPERIMENT II

Having demonstrated the rapid acquisition of task T by the two groups trained on the complex geometric tasks, the question arose as to how far the concepts which the children were assumed to be using might generalize to new material.

Method

Shift tasks

Two new sets of pictures based on the same categories used in section I were devised.

T_1 was a set of photographs in which each example of the categories was the same in nature as that used in task T, but differed in quality (e.g. the whole figure of a cat on a white background was replaced by the head of a cat on a dark background; a mongrel dog facing the camera was replaced by a side view of a poodle: sandals on a white background were replaced by tennis shoes on a dark background, and so on). The stimulus pictures which had been a baby, a pig, a cup, a sofa and a dress, were in this situation: a man's torso, a horse, a plate, a bookcase and a hat.

T_2 was a set of outline drawings on a plain background, most of which were new examples of the same five categories. The stimulus cards were: a crocodile, a baby's cot, a little girl, a ski-jumper and a tea-pot. The response cards were: shoe, woman's torso, desk, jug, kangaroo, boy's head, dressing-table, egg-cup, tie, eagle, knife, T.V. set, deer, trousers, Indian fakir, armchair, hat, elephant's head and forelegs, glass and a man.

Subjects

By the time this part of the experiment had been reached, one subject had been lost from each of the groups III, I and control, and two subjects from group II. All those mentioned earlier as unusual for their group remained, however, so the three experimental groups still retained the original low scorers and the control group the fast learner. The numbers of subjects available were, therefore, as follows: group I, 7; group II, 7; group III, 8; control, 12.

Procedure

All the children were given two trials on T_1 and one trial on T_2, using the same procedure as formerly. No additional instructions were used.

Results

The results are summarized in Table 5. Although all four groups showed a decrease in score of between 2½ and 3 points on these three trials compared with the last four trials of task T, the significant differentials among the groups were maintained; groups III and II were again significantly superior to groups I and the control group; and group III non-significantly higher than group II.

Table 5. *Analysis of score data on three shift trials (two trials on* T_1 *and one trial on* T_2*)*

Analysis of variance

Source	S.S.	D.F.	M.S.	F	P
Between groups	344·99	3	115·00	11·09	< ·001
Residual	1016·59	98	10·37	—	—

Tukey comparison of means

Group	III	II	Control	I
Means	15·83	14·90	11·86	11·57
III	—	N.S.	< 0·01	< 0·01
II	—	—	< 0·01	< 0·01
Control	—	—	—	N.S.

These data do not at first seem to add substantially to what was already known about the groups, beyond indicating that the group differentials were sustained after the shift to new material. This finding suggests that members of groups II and III (fairly difficult and difficult training, respectively) had sufficient familiarity with the five concepts to operate at a fairly high level even in what was, for them, a taxing situation in which they had to identify new pictures and classify them with new stimulus objects.

Closer inspection of the scores, however, revealed differences in the behaviour of subjects within the groups which cast further light on the acquisition of the concepts under investigation. Although some members of all four groups had obtained high scores (18, 19 or 20 out of a possible maximum of 20) during training on T, it was evident that those few from groups I and control who did so had lost more on shift to the new material than those from groups II and III, who with two exceptions had attained this criterion. It thus seemed that efficiency on the shift tasks might be due not only to the level attained on T but also to the number of trials of practice at a high level. The number of trials taken to reach a score of 18 or above was therefore correlated with average score on the three shift trials. In view of the skewed distribution of scores on the former variable, Kendall's rank correlation was computed, giving $\tau = -0{\cdot}644$ ($z = 5{\cdot}37$; $P < 0{\cdot}00006$, two-tailed). It thus appeared that the earlier a subject obtained a high score on T, the better was his performance on transfer to the parallel tasks. This evidence indicates that a degree of overlearning on T facilitated the use of the categories in dealing with new material.

The results accord with the theory of cognitive development proposed by Mandler (1962) who sought to reconcile associationist and cognitive or 'structural' theories of learning, by proposing that cognitive structures develop out of stimulus–response associations during overlearning. At this stage symbolic analogues, or analogic structures, replace the simpler associations. In elaborating his theory, Mandler reviewed evidence on the 'warm-up' effect, learning sets and the relation between overlearning and transfer. Drawing on experiments in the fields of both animal and human learning, he has pointed to 'the similarity of the general function of the relationship between transfer and degree of prior learning'.

The great majority of subjects showed continuous learning, and an orderly reduction in errors; some (mostly from groups II and III) attained virtually error-free performance fairly early during acquisition of *T* and thereafter maintained this high level. The correlation between rate of initial acquisition and level of score on transfer to the shift tasks suggests that during the later trials on *T*, those subjects who had attained a high level early developed at least rudimentary concepts or structures for the five categories; these they subsequently applied with considerable ease when presented with new stimulus material. The precise nature of these symbolic analogues is unknown, since the subjects indulged in minimal spontaneous verbalization, and it is doubtful whether children of this age would be capable of a valid verbal translation of their own conceptual processes.

Conclusions

The level attained by the subjects when shifted to parallel versions of the transfer task was found to be directly related to the rapidity of acquisition of *T*, and thus to the amount of overlearning. These in turn were of course strongly influenced by the complexity of intervening training.

III. TRANSFER BETWEEN DISSIMILAR TASKS: ITS RELATION TO OVERLEARNING: EXPERIMENT III

Summarizing the results so far, it seems clear that the learning of tasks of the kind used is greatly facilitated by previous training on complex categorizing tasks of widely different content. Nevertheless, both in the experiment with older, imbecile children, and in this investigation no significant difference was found in rate of acquisition of *T* between two groups, one of which was trained on a difficult task and the other on a fairly difficult one, though most of the results were in the expected direction. To this extent the hypothesis was not confirmed, even though in the present investigation the differential difficulty of the two tasks had been made greater than in the experiment with imbeciles. Two hypotheses were considered that might account for this finding: (1) that the threshold of learning for transfer is fairly low, a small amount of complex training being all that is required to produce subsequent facilitation on a different task, and any training over and above this level is therefore redundant; (2) that the performance of group II (fairly difficult training) on *T* equalled that of group III (difficult training) partly by virtue of an overlearning effect. In both investigations the number of trials on the intervening tasks had been held constant for the training groups, and hence, inevitably, those trained on the less difficult of the two more complex tasks (task II) had reached their asymptote earlier than those trained on the more difficult task (III), and thereafter had 'overlearned'.

The role of overlearning in relation to positive and negative transfer has been explored in a variety of situations, but in none so systematically as in reversal behaviour in lower animals. Here the overlearning of a discrimination has been shown by several investigations to facilitate reversal learning (Reid, 1953; for recent reviews, see Paul, 1965; Mackintosh, 1965*b*), although it has also been shown to impede an extra-dimensional shift (Mackintosh, 1964). A series of recent studies by Youniss & Furth (1964*a*, *b*, 1965) and by Furth & Youniss (1964) has been concerned with the effects of overtraining on reversal and non-reversal shifts in children aged 6–8 years. Among other things they found that overtraining facilitated reversal of a spatial discrimination only in the presence of irrelevant cues. Although the tasks used in the present experiments were widely different from those employed in studies of reversal shifts, it seemed possible that the phenomena might be related.

Method

With a fresh intake of children into the local schools, it was possible to select a group of eleven, matched with the previous four groups on the relevant variables: age, score on T, errors and ratio of girls to boys. These children (now termed group V) were trained on task II (fairly difficult) to a criterion of 14 or 15 first correct placements out of the maximum score of 15 (i.e. to asymptote with no over-training), and then transferred to task T for ten trials. If the hypothesis concerning a threshold effect were correct, this group should be identical to groups II and III in Expts I and II. If, however, the overlearning hypothesis were correct, this group should lie below group II but above the control group in its acquisition of T.

Results

Table 6 gives the initial data on the new group V and a comparison of their pretest/retest difference scores with those of groups II and control. Unlike the previous four groups, the range of pretest/retest scores for the new group was very great; and, because the variances were not homogeneous, a non-parametric test was used for comparisons in this one instance.

Table 6. *Data on new group, V, trained to criterion only on task II, and comparison of pretest/retest differences with groups II and control* (*cf. Table* 1)

n...	11	Comparisons with groups II, III and control
Mean age (months)	50·09	$F < 1$ in all cases
Mean pretest score, task T	6·73	
Mean pretest errors, task T	31·00	
No. of boys	7	

	Pretest/retest differences		
	Group II	Group V	Control
Means	5·11	4·09	2·08
S.D.	2·37	4·37	2·10
n	9	11	13

Kruskall-Wallis one-way analysis of variance (Siegel, 1956): $H = 106{\cdot}176$ ($P < 0{\cdot}001$)
Mann–Whitney tests: group II and group V ($U = 36$; N.S.); group V and control ($U = 57{\cdot}5$; N.S.)

Inspection of the difference scores for the new group suggested a correlation between number of trials to criterion on the intervening training task and amount of gain on immediate transfer to T. The rank correlation was $\rho = 0{\cdot}646$ ($t = 2{\cdot}537$; D.F. 9; $P < 0{\cdot}05$). Thus, it appeared that the more trials a subject required to reach the pre-set criterion, the greater the gain on immediate retest.

Table 7 gives the results of comparisons of the new group V with groups II, III and control on the ten acquisition trials for *T* and on the three shift trials. It appears from the results that group V is significantly inferior to groups II and III in acquisition of task *T* and in performance after the shift to new material; the 'overlearning' hypothesis' is thus supported.

Table 7. *Comparisons of group V with groups II, III and control*

(*a*) Analysis of variance on scores over 10 trials on T

Source	S.S.	D.F.	M.S.	*F*	*P*
Between groups	1014·29	3	338·10	21·94	< 0·001
Residual	6409·46	416	15·41	—	—

Tukey comparison of means

Group	III	II	V	Control
Means	16·57	16·11	14·55	12·70
III	—	N.S.	< 0·01	< 0·01
II	—	—	< 0·01	< 0·01
V	—	—	—	< 0·01

(*b*) Analysis of variance on scores on 3 shift trials

Source	S.S.	D.F.	M.S.	*F*	*P*
Between groups	279·89	3	93·30	8·06	< 0·001
Residual	1274·18	110	11·58	—	—

Tukey comparison of means

Group	III	II	V	Control
Means	15·83	14·90	12·91	11·86
III	—	N.S.	< 0·05	< 0·01
II	—	—	< 0·10	< 0·05
V	—	—	—	N.S.

Conclusions

The data suggest that rapid acquisition of a task of this kind is facilitated either by learning a difficult task, not necessarily to perfection, or by learning and considerable overlearning of a fairly difficult task. Further work will, however, be needed to clarify the issue, since, as already mentioned, a detailed analysis of the learning of each child on these tasks indicated a hierarchical order of difficulty of the categories. Thus, during the learning of the quadrilateral sorting task by group III some overlearning of the easier categories had inevitably occurred while the more difficult categories were being acquired. As yet, therefore, the relative contributions of complexity and overlearning of difficult discriminations to subsequent transfer are not established; both are probably important in any learning situation of this kind at this age. However, the data from group I suggest that overlearning of an easy discrimination may initially impede performance on a difficult transfer task.

A further hypothesis to be tested is that a potential difference between groups II and III after pretraining was masked by the level of difficulty of task *T*. In other words, if *T* were made more difficult, subjects trained on the quadrilateral sorting task might show more rapid acquisition than subjects overtrained on the less difficult task.

DISCUSSION

It is not as yet entirely clear how these findings may be interpreted in the wider context of general theories of transfer of learning, since the data on which such theories are based come predominantly from experiments using tasks or subjects differing considerably from those considered in this paper. Hence the discussion will be confined to a brief account of the factors which are thought to underly the observed behaviour, and to a possible relation with some recent findings in the field of animal learning.

The results suggest that, at least for naïve subjects, a set or ability to categorize may be developed in one learning situation and transferred to a widely different problem. Furthermore, transfer is related both to the complexity of the training task and also to the degree of its overlearning. These findings have emerged from studies using different populations, different techniques and different material. Moreover, although transfer occurs when there are identical elements or principles common to two situations, it is now evident that it can also take place across tasks where the only similarity lies in the need to categorize to achieve correct solutions. Practice in categorization appears to improve the process of categorization in a different situation, perhaps by means of increased attention to the *details* of the new task and of experience in making common responses to dissimilar stimuli (Clarke *et al.* 1966).

It was impossible to know how many partly formed responses to this material the children came with, and to what extent they acquired new systems of classification. In view of the fairly systematic emergence of correct responses to categories in a hierarchical order it seems reasonable to suppose that the subjects probably already had some idea of at least some of the categories (e.g. animals and tableware), which were then fairly rapidly consolidated in the experimental situation. The less familiar categories might then have been built up by contrast. Those who could rapidly acquire a high level of performance of T were probably able, during their period of overtraining at this level, to develop some undefined symbolic analogues or concepts which made it possible for them to sort different material in the same categories fairly efficiently. The high correlation between degree of overlearning on T and performance on the shift tasks indicates that the more the overlearning of this material, the better the development of the symbolic analogue or concept needed for efficient performance in a similar situation.

Although the precise relations between complexity and transfer need further clarification, the tentative conclusion may be drawn that comparisons of learning and transfer between subjects of different ages or different abilities are not necessarily to be taken at their face-value unless some experimental adjustment of task complexity is incorporated within the design. The same point may apply to inter-species comparisons.

Some confirmation of this suggestion appears in results obtained by Mackintosh (1965*a*) in connexion with the demonstration of an overtraining reversal effect in rats and chicks. Mackintosh examined an apparent discrepancy between the behaviour of rats and chicks found by Brookshire, Warren & Ball (1961). They found the well-established positive effect of overlearning on the reversal of a brightness discrimination in rats, but their chicks learned the same reversal more slowly after overtraining than non-overtrained birds. Mackintosh repeated the experiment with similar results, but went on to show that overtraining significantly facilitated reversal learning for chicks, which have markedly better vision than rats, if they were trained on a discrimination of appropriate difficulty *for them*, in this case an orientation discrimination involving two irrelevant cues. Mackintosh made his prediction on the basis of an explanation of the cause of the overtraining reversal effect proposed by Sutherland (1959) and himself (Mackintosh, 1962) which is directly applicable to the specific behaviour observed in lower animals in a reversal situation and cannot easily be extended to accommodate the results of the present experiment. Nevertheless, it seems possible that the phenomena are related, and that in comparative studies of learning and transfer less emphasis should perhaps be put on the absolute properties of stimuli and responses, and more attention should be given to the relative properties of the experimental situation in terms of level of difficulty for the subjects' age, ability, place in the phylogenetic scale, and previous relevant experiences.

The role of learning transfer, and of the environmental variables inducing it, may well be important in cognitive development, by extending the organism's repertoire of behaviour and raising him above a narrow dependence upon specific learning. It is surprising that so little research has been carried out in this area, which appears to be a rewarding one, in spite of the clear lead from Hebb (1949) who wrote: 'If the learning we know and can study, in the mature animal, is heavily loaded with transfer effects, what are the properties of the original learning from which those effects came? How can it be possible even to consider making a theory of learning in general from the data of maturity alone? There must be a serious risk that what seems to be learning is really half transfer. We cannot assume that we know what learning transfers and what does not: for our knowledge of the extent of transfer is also derived from behaviour at maturity, and the transfer from infant experiences may be much greater and more generalized' (p. 110).

The writers gratefully acknowledge the provision of research facilities by Mr S. W. Hobson, Chief Education Officer, Kingston-upon-Hull; the co-operation of staff and children of Macmillan House School and Northumberland Avenue School; the assistance of the University Photographer, Mr A. Marshall, and of Mr J. Devine, Chief Technician and his staff; helpful advice from Mr A. B. Royse and Professor A. D. B. Clarke.

The investigation was made possible by a generous grant from the Nuffield Foundation.

REFERENCES

BROOKSHIRE, K. H., WARREN, J. M. & BALL, G. G. (1961). Reversal and transfer learning following overtraining in rat and chicken. *J. comp. physiol. Psychol.* **54**, 98–102.

CLARKE, A. D. B. & COOPER, G. M. (1966). Age and perceptual-motor transfer in imbeciles: task complexity as a variable. *Brit. J. Psychol.* **57**, 113–19.

CLARKE, A. M., COOPER, G. M. & HENNEY, A. S. (1966). Width of transfer and task complexity in the conceptual learning of imbeciles. *Brit. J. Psychol.* **57**, 121–8.

FURTH, H. G. & YOUNISS, J. (1964). Effect of overtraining on three discrimination shifts in children. *J. comp. physiol. Psychol.* **57**, 290–3.

HEBB, D. O. (1949). *The Organization of Behavior*. New York: Wiley.

HELSON, H. (1959). Adaptation level theory. In *Psychology: a Study of a Science*, vol. I, ed. S. Koch, pp. 565–621. New York: McGraw-Hill.

MACKINTOSH, N. J. (1962). The effects of overtraining on a reversal and a non-reversal shift. *J. comp. physiol. Psychol.* **55**, 555–9.

Mackintosh, N. J. (1964). Overtraining and transfer within and between dimensions in the rat. *Quart. J. exp. Psychol.* **16**, 250–6.

Mackintosh, N. J. (1965*a*). Overtraining, reversal, and extinction, in rats and chicks. *J. comp. physiol. Psychol.* **59**, 31–6.

Mackintosh, N. J. (1965*b*). Selective attention in animal discrimination learning. *Psychol. Bull.* **64**, 124–50.

Mandler, G. (1962). From association to structure. *Psychol. Rev.* **69**, 415–27.

Paul, C. (1965). Effects of overlearning upon single habit reversal in rats. *Psychol. Bull.* **63**, 65–72.

Reid, L. S. (1953). The development of noncontinuity behaviour through continuity learning. *J. exp. Psychol.* **46**, 107–12.

Ryan, T. A. (1959). Multiple comparisons in psychological research. *Psychol. Bull.* **56**, 26–47.

Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

Sutherland, N. S. (1959). Stimulus analysing mechanisms. In *Proceedings of a Symposium on the Mechanization of Thought Processes*, vol. II, 575–609. London: H.M.S.O.

Youniss, J. & Furth, H. G. (1964*a*). Reversal learning in children as a function of overtraining and delayed transfer. *J. comp. physiol. Psychol.* **57**, 155–7.

Youniss, J. & Furth, H. G. (1964*b*). Reversal performance in children as a function of overtraining and response conditions. *J. exp. child Psychol.* **1**, 182–8.

Youniss, J. & Furth, H. G. (1965). Discrimination shifts as a function of degree of training in children. *J. exp. Psychol.* **70**, 424–7.

(*Manuscript received* 6 *November* 1965)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 375–380
Printed in Great Britain

AVOIDANCE RESPONSES AS DISCRIMINATED OPERANTS

By J. D. KEEHN
Addiction Research Foundation, Toronto 5, Ontario

In response to a comment by Brown (1965) it is argued that stimuli associated with electric shock in avoidance-learning experiments do not necessarily acquire an attribute of aversiveness but function as discriminative stimuli. The usual assumption that avoidance behaviour is not reinforced by shock omission but is acquired coincidentally through its collateral termination of a conditioned aversive state of affairs is rejected in the light of experimental evidence.

I. Introduction

A recent comment by Brown (1965) may be summarized as follows: on the basis of experimental evidence Keehn (1959*a*) arrived at conclusion A; in the light of compatible evidence Brown reached a different conclusion, B. If an aversive stimulus is defined as p then Keehn is correct, but if it is defined as q then Brown is correct. Brown prefers definition q; therefore Keehn is wrong.

Conclusions A and B pertain to the functions of stimuli in experiments on avoidance learning. The writer regards them as discriminative in the operant sense of setting the occasion for a response (Skinner, 1938); Brown considers them to be aversive. Since Brown acknowledges that 'our theoretical concepts are sometimes shaped by the experimental situations we adopt', the writer proposes first to demonstrate that his experimental arrangement is superior to the traditional one that Brown prefers, then to consider some difficulties in defining a stimulus as aversive on the basis of data generated by this technique, and finally to examine some implications of an operant analysis of avoidance responding.

II. Studying avoidance behaviour

Many different experimental arrangements have been used to study avoidance behaviour, most of them involving groups of animals trained by the discrete trial method (Solomon & Brush, 1956). One of the most popular and convenient of these arrangements is the double compartment box described by Brown. An animal is placed in one compartment and after a time a signal is presented. If the animal quickly runs to the other side the signal is terminated and shock is avoided; if not, shock is delivered until the animal escapes both signal and shock by crossing compartments. After an inter-trial interval the procedure is repeated on the other side until eventually the animal shuttles back and forth according to the onset of the signal in each compartment. The data generated by this method are usually plotted as learning curves showing the increasing number of responses made to the signal (correct responses) as training proceeds, or as changes in response latency to the signal. Such data suggest an acquired property of the signal to evoke an avoidance response. The signal is usually referred to as a *conditioned stimulus*, and although it was once thought that the signal evoked the avoidance response directly the more common theoretical interpretation now is that fear is classically conditioned to the signal, and that the avoidance response is reinforced by fear reduction consequent on the termination of the signal.

Characteristically, however, after a few shocks animals begin shuttling between compartments without waiting for the signal (e.g. May, 1948; Kamin, 1954; Brown, 1965). Eventually inter-trial responses extinguish and the terminal behaviour is as described above, but the approach to this performance, *when all the data are considered*, is not from no response to response to the signal, but from over-responding to responding only in the presence of the signal. This suggests a different interpretation of avoidance behaviour from the one given above: the appearance is of *emitted* operant behaviour coming under the control of a *discriminative stimulus* (Skinner, 1938).

A free operant avoidance training procedure was invented by Sidman (1953). In the Sidman procedure brief shocks are scheduled to occur at regular intervals (shock-shock, or SS interval) unless an avoidance response is made in the meantime, in which case the scheduled shock is postponed for a specified time (response-shock, or RS interval). When no exteroceptive stimulus is given the procedure is called *non-discriminated* avoidance, but *a discriminated* procedure is possible if a signal is presented during the SS and RS intervals. These intervals may then be designated SL; LS and RL; LS respectively, where SL is the time from an unavoided shock to signal onset, LS the time from signal onset to the next programmed shock, and RL the time that a response postpones signal onset.

Compared with the conventional procedure LS is the interstimulus interval and SL the inter-trial interval. (LS is usually called the CS–UCS interval, even by writers who argue that avoidance behaviour is instrumentally reinforced. The terminology is inappropriate because avoidance responses are not conditioned responses in the Pavlovian sense; they are unconditioned responses to shock only under very special circumstances.) There is no RL interval in the conventional procedure because inter-trial responses are either prevented or are non-functional. Such responses are regarded as a nuisance in the conventional avoidance procedure and are not subjected to systematic analysis even when investigators unwittingly make them functional by postponing trials until they cease. The writer regards the Sidman procedure as superior to the traditional methods of studying avoidance behaviour because it takes account of all of the data, because the traditional technique is a special case of Sidman's discriminated method (in which interval responses are non-functional), and because Sidman's technique utilizes the direct behavioural data of single animals rather than averaged behaviour of groups. With Sidman's technique it is possible to generate all of the data recorded by conventional means, and many more besides. These extra data free the theorist from limitations imposed by the information available from discrete trial avoidance techniques.

III. DEFINING AVERSIVENESS

Stimuli may be defined as aversive, or appetitive, either intuitively or according to the behaviour they control. Intuitive definitions are usually applied to so-called primary reinforcers, like food or electric shock, and behavioural definitions to stimuli that come to control behaviour through certain experimental manipulations. This is an uncomfortable dichotomy and behavioural definitions may be used in every case, although this too generates some intuitive discomfort as is considered below.

Behaviourally a stimulus is defined as aversive (*a*) if behaviour that terminates it is maintained, (*b*) if behaviour that avoids it is maintained, (*c*) if behaviour that produces it is suppressed; or in some other ways (Leitenberg, 1965). Usually electric shock satisfies all these criteria and may be regarded as aversive on this account, in accordance with intuition. Intuition demands that a stimulus paired with electric shock also becomes aversive, but according to the experimental method of pairing behavourial criteria of aversiveness may be exhibited or they may not.

An experimental paradigm (Keehn, 1959*a*) that does not demonstrate aversiveness in a pre-shock signal according to criteria (*b*) and (*c*) may be symbolized $(\mathrm{RL}x)^{\mathrm{D}} \rightarrow (\mathrm{LS}y)^{\mathrm{L}}\mathrm{R} \rightarrow (\mathrm{RL}x)^{\mathrm{D}} \ldots$, meaning that each response in stimulus D postpones stimulus L for x seconds, and that in stimulus L a shock follows in y seconds unless a response occurs in the interval, and that each response in L terminates the stimulus (and shock when it occurs) and reinstates the $(\mathrm{RL}x)^{\mathrm{D}}$ condition. In these circumstances responses congregate in the LS interval. Stimulus L is not avoided and so is not aversive by criterion (*b*); it is terminated so it is aversive by criterion (*a*). But criterion (*a*) may only be satisfied coincidentally because the response that terminates it is necessary to avoid shock, so there is no unambiguous evidence that the stimulus is aversive. The usual behaviouristic interpretation of avoidance learning assumes that shock avoidance is coincidental upon stimulus termination, but this may be a gratuitous assumption (see below).

A slightly different experimental arrangement removes the ambiguity. This arrangement (Sidman, 1957) may be symbolized $(\mathrm{RL}x)^{\mathrm{D}} \rightarrow (\mathrm{RS}y)^{\mathrm{L}}.\mathrm{S} \rightarrow (\mathrm{RL}x)^{\mathrm{D}} \ldots$, meaning that each response in stimulus D postpones stimulus L for x seconds, and that every response in stimulus L postpones shock for y seconds, and that a shock terminates the stimulus and reinstates the $(\mathrm{RL}x)^{\mathrm{D}}$ condition. Once again stimulus L is not avoided and so is not aversive by criterion (*b*); nor is it terminated so it is not aversive by criterion (*a*) either. But again there is an ambiguity because the stimulus is only terminated on the occurrence of a shock. Shocks are avoided in the presence of the stimulus even though avoiding the shock does not terminate the signal. However this performance is only maintained if y is greater than x. When x is greater than y responses congregate in the RL interval. Stimulus L is avoided, and so is aversive by criterion (*b*), and also by criterion (*a*) because the stimulus is allowed to terminate whenever it occurs. As stimulus termination is accompanied by shock this means that the stimulus is *more aversive than shock* (Sidman & Boren, 1957) because the stimulus is avoided but shock is not.

Slight differences in experimental conditions, then, generate behaviour according to which a stimulus may be defined as aversive or not. This is one reason why the writer does not believe that a stimulus acquires an attribute of aversiveness through association with electric shock, and prefers to explore its discriminative properties instead. Even electric shock does not exhibit the criteria of aversiveness if it is used as a discriminative stimulus for positive reinforcement (Holz & Azrin, 1961). Conversely a discriminative stimulus for positive reinforcement may exhibit aversive properties by criterion (*a*), Keehn (1962). Because a stimulus is not aversive it does not follow that it cannot reinforce. Brown seems to think otherwise: '...it is perfectly justifiable to claim that presence-of-signal and absence-of-signal act as a pair of discriminative stimuli, but to deny that these stimuli may also possess reinforcing properties would seem to be unwarranted.' The operations for establishing discriminative stimuli and conditioned reinforcers are identical, but the writer has reservations about the utility of the concept of conditioned reinforcement (Keehn, 1962), and he has indicated the possible *positive* reinforcing role of pre-shock signals for observing behaviour (Keehn, 1959*a*).

IV. STIMULUS CONTROL OF AVOIDANCE BEHAVIOUR

Discriminative stimulus control over an emitted operant is gained when the response is reinforced in the presence of a stimulus (S^D) but not in its absence (S^Δ). Because behaviour that terminates or avoids it increases in strength, electric shock is defined as a negative reinforcer. In the conventional avoidance experiment responses in the presence of the pre-shock signal terminate or avoid shock, but responses in the absence of the signal have no effect. Conditions are appropriate, then, for the signal to be established as S^D and its absence as S^Δ, and experimental data support this interpretation: avoidance responses come to concentrate in the presence of the signal (S^D) and extinguish in its absence (S^Δ).

The analogous situation with positive reinforcement may be easier to understand because it is more tangible; in S^Δ a response produces no food and extinguishes, in S^D food is procured and the response is strengthened. The experimenter is in a favourable position because he can see the reinforcer, an advantage he seems to lose in the case of avoided shocks. But the difference is illusory, because the experimenter is just as aware that a shock has been avoided as he is that a response has been followed by food, or that unnecessary behaviour has been emitted in either case. The experimenter's operations are the same in the positive and negative cases, even literally, for he may press a button to set up positive reinforcement on the one hand or to postpone shock on the other, and this behaviour is occasioned by the behaviour of the rat, which may also literally be the same in either case. Difficulties arise because in the negative case the experimenter shifts from an operational to an intuitive stance. He denies that avoidance behaviour is reinforced, only escape behaviour that coincidentally avoids, because he intuits that the laboratory rat cannot discriminate that electric shock is imminent, or that it has been avoided. Even if intuition of this kind were scientifically permissible it is doubtful if the conclusion is correct in the present case. The laboratory rat is capable of far more elaborate temporal discriminations than those required in the conventional avoidance experimental arrangement (Keehn, 1965). There are no experimental reasons for denying that laboratory animals can anticipate shock and therefore no justification for departing from a behavioural definition of pre-shock signals as discriminative stimuli established through differential (negative) reinforcement.

V. ACQUISITION OF AVOIDANCE BEHAVIOUR

The above account of the discriminative stimulus control of avoidance behaviour presupposes that avoidance responses occur. An operant must be emitted before it can be reinforced, so some discussion on how avoidance responses come to be emitted in the first place may be in order. From a strictly Pavlovian standpoint there is no problem, for an avoidance response is merely an antedating escape response that is evoked by a conditioned stimulus instead of unconditioned shock. But Pavlovian conditioning cannot account for avoidance learning (Mowrer & Keehn, 1958; Keehn, 1959*a*, *b*). There is also no special difficulty with more complex conditioned drive theories for they too stress the conditioned evoking powers acquired by pre-shock stimuli. However, operant responses are emitted not evoked, and until they are emitted differential reinforcement cannot be applied, so the investigator must either find some way of shaping avoidance responding, or wait until responses happen to occur. When an avoidance response is emitted the only controlling operation available to the experimenter is withholding shock (negative reinforcement), and this is only meaningful in a context in which shocks normally occur in the absence of avoidance responses. Hence before negative reinforcement can be applied a condition must be created in which shocks predictably recur. This can be done either by presenting shocks at fixed regular intervals or by presenting a signal prior to irregularly spaced shocks. The latter is the method usually adopted in conventional avoidance procedures with the added contingency that an escape response terminates both signal and shock. This serves to strengthen the *escape* response, and if escape and avoidance responses are topographically the same, as they usually are, avoidance responses thereby become more likely to be emitted and so available for reinforcement by shock omission.

It is not necessary for escape and avoidance responses to be alike. In one experiment (Keehn, 1959*b*) an avoidance response was a forward run in an activity wheel, but when shock occurred it was terminated only when the animal turned about. The experiment demonstrated independent sources of reinforcement for escape and avoidance responses and that avoidance is not antedating escape, but did not settle the question of how avoidance responses are reinforced—either by shock avoidance or escape from aversive stimulation. The experimental outcome probably depended on the fortuitous choice of escape and avoidance responses, for had running been the escape response and turning about the avoidance response it is likely that avoidance learning would not so easily have occurred. By chance the avoidance response was one that is emitted with high probability when rats are shocked in activity wheels, even when running is not functional (Mowrer & Keehn, 1958, control animal). When other responses are chosen for study avoidance behaviour is much more difficult to establish (Keehn & Sabbagh, 1958; Meyer, Cho & Wesemann, 1960). Conventional avoidance theory makes no provision for this differential for it draws its data from experimental arrangements in which avoidance responses are frequently emitted. Operant theory is prepared for the difficulty because it emphasizes the requirement of response emission before reinforcement and recognizes that avoidance responses are difficult to shape. The objective of the student of operant behaviour is to devise means of ensuring that the behaviour he wishes to control occurs. What is the theorist to do when behaviour he believes he can adequately explain fails to occur?

VI. CONCLUSION

It has been argued that the free operant (Sidman) technique of studying avoidance behaviour is superior to a methodology of discrete trials because it systematically generates data that occur 'spontaneously' in other methods, and because the discrete trial procedure is a special case of the Sidman method. Data gathered by this technique have been cited which show that stimuli paired with electric shock do not unambiguously exhibit aversive properties, and it has been argued that they can consistently be regarded as discriminative stimuli in the operant sense. The arguments rest on the assumption that avoidance of shock can be reinforcing in itself, and on some evidence running counter to the usual supposition that avoidance behaviour is only a coincidental accompaniment of escape from conditioned aversive stimulation.

REFERENCES

BROWN, R. (1965). Discrimination of avoidable and unavoidable shock. *Brit. J. Psychol.* **56**, 275–83.

HOLZ, W. C. & AZRIN, N. H. (1961). Discriminative properties of punishment. *J. exp. Anal. Behav.* **4**, 225–32.

KAMIN, L. J. (1954). Traumatic avoidance learning: the effects of CS–US interval with a trace conditioning procedure. *J. comp. physiol. Psychol.* **47**, 65–72.

KEEHN, J. D. (1959*a*). The effect of a warning signal on unrestricted avoidance behaviour. *Brit. J. Psychol.* **50**, 125–35.

KEEHN, J. D. (1959*b*). On the non-classical nature of avoidance behavior. *Amer. J. Psychol.* **72**, 243–7.

KEEHN, J. D. (1962). The effect of post-stimulus conditions on the secondary reinforcing power of a stimulus. *J. comp. physiol. Psychol.* **55**, 22–6.

KEEHN, J. D. (1965). Temporal alternation in the white rat? *J. exp. Anal. Behav.* **8**, 161–8.

KEEHN, J. D. & SABBAGH, U. (1958). Conditioned inhibition and avoidance learning. *Psychol. Rep.* **4**, 547–52.

LEITENBERG, H. (1965). Is time out from positive reinforcement an aversive event? A review of the experimental evidence. *Psychol. Bull.* **64**, 428–41.

MAY, M. A. (1948). Experimentally acquired drives. *J. exp. Psychol.* **38**, 66–77.

MEYER, D. R., CHO, C. & WESEMANN, A. S. (1960). On problems of conditioning discriminated lever-press avoidance responses. *Psychol. Rev.* **67**, 224–8.

MOWRER, O. H. & KEEHN, J. D. (1958). How are intertrial 'avoidance' responses reinforced? *Psychol. Rev.* **65**, 209–21.

SIDMAN, M. (1953). Avoidance conditioning with brief shock and no exteroceptive warning signal. *Science* **118**, 157–8.

SIDMAN, M. (1957). Conditioned reinforcing and aversive stimuli in an avoidance situation. *Trans. N.Y. Acad. Sci.* ser. ii, **19**, 534–44.

SIDMAN, M. & BOREN, J. J. (1957). The relative aversiveness of warning stimuli and shock in an avoidance situation. *J. abnorm. soc. Psychol.* **55**, 339–44.

SKINNER, B. F. (1938). *The Behavior of Organisms*. New York: Appleton-Century-Crofts.

SOLOMON, R. L. & BRUSH, E. S. (1956). Experimentally derived conceptions of anxiety and aversion. In *Nebraska Symposium on Motivation* ed. M. R. Jones. Lincoln: University of Nebraska Press.

(*Manuscript received* 22 *March* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 381–389
Printed in Great Britain

THE EFFECTS OF NOISE AND DOUBLING THE SIGNAL FREQUENCY ON INDIVIDUAL DIFFERENCES IN VISUAL VIGILANCE PERFORMANCE

By D. R. DAVIES and G. R. J. HOCKEY*
Department of Psychology, University of Leicester

Four groups of 6 extraverts and 4 groups of 6 introverts, selected by the Maudsley Personality Inventory, performed a 32 min visual cancellation task under one of two conditions of signal frequency, high and low, and in either noise (95 db) or quiet (70 db). In quiet, at both levels of signal frequency, extraverts showed a steady decline in the number of signals detected correctly but introverts did not. Neither group showed a decrement under noise conditions. Noise, compared with quiet, significantly increased the number of correct detections made by extraverts under low signal frequency conditions but a similar increase under high signal frequency conditions was not significant. The addition of noise had no significant effect on the number of correct detections made by introverts. Doubling the signal frequency had no significant effect on the performance of introverts or extraverts in either noise or quiet. Introverts made significantly more errors of commission in quiet than in noise while extraverts made significantly more in noise than in quiet. Possible reasons for the findings are discussed.

Many experiments have demonstrated that, with increasing arousal, performance at several different tasks improves up to an optimal point and thereafter declines (Hebb, 1955; Duffy, 1957, 1962; Malmo, 1959). The relation between arousal level and performance is thus considered to take the form of an inverted U-shaped curve. A subject's arousal level is sometimes regarded as lying on a continuum bounded at one end by deep sleep and by extreme excitement at the other. Various physiological measures, such as EEG frequency and amplitude, heart rate and skin conductance have been thought to provide an indication of a subject's arousal level (Duffy, 1962) but some investigators have preferred to rely on operational definitions of arousal, such as 'the inverse probability of falling asleep' (Corcoran, 1963). Corcoran considered sleep-deprived subjects to be at a lower level of arousal than subjects who have slept normally; quite apart from the fact that this accords well with common sense, there is some physiological evidence to support the assumption (Corcoran, 1964). Intense stimulation, in the form of white noise for example, is generally held to raise the arousal level. Given the assumption of an inverted U-relation between arousal and performance, the effect on performance of adding noise to a task situation should depend upon the initial arousal level of the subject. A greater improvement in performance would be expected from subjects who are initially at a low level of arousal, for example, sleep-deprived subjects, than from subjects who have had their normal night's sleep, who are more highly aroused and thus nearer the optimal arousal point to begin with. This expectation has been confirmed by Corcoran (1962*a*).

The amount of sensory input can also be increased by increasing the quantity of relevant information fed to the subject in a given interval of time, for example, increasing the signal frequency in a vigilance task. Thus the effects of increasing the signal frequency on the performance of sleep-deprived and 'normal' subjects should parallel the effects of added noise. This has also been confirmed by Corcoran (1962*b*). Both noise and signal frequency seem to be arousing in their effects, therefore, given that sleep-deprived subjects are less aroused than subjects who have slept normally.

---

* Now at M.R.C. Applied Psychology Research Unit, Cambridge.

The performance of extraverts and introverts, whether selected by the Heron Sociability Scale (Heron, 1956) or by the Maudsley Personality Inventory (Eysenck, 1959) would appear to be analogous to that of sleep-deprived and 'normal' subjects and it is possible to think of introverts as being more aroused than extraverts (Broadbent, 1963; Corcoran, 1965). Bakan (1959) found that the addition of a secondary task, which had the effect of increasing signal frequency, resulted in a significantly greater improvement in the vigilance performance of extraverts as compared to introverts.

Two assumptions underlay the present experiment. They are: (*a*) that noise and increasing the signal frequency are arousing in their effects, and (*b*) that introverts are more aroused than extraverts. The purpose of the experiment was to observe the effects of adding noise and doubling the signal frequency on the vigilance performance of introverts and extraverts. Expectations were: (i) that in quiet the vigilance performance of introverts will be superior to that of extraverts, (ii) that extraverts will show a greater improvement when noise is added and that this improvement will be greater under conditions of low signal frequency, and (iii) that compared to quiet conditions the performance of introverts will either show no change or be impaired under conditions of noise, the probability of impairment being greater under conditions of high signal frequency.

METHOD

Subjects, divided into groups of introverts and extraverts, performed a visual cancellation task lasting for 32 min in which they were instructed to detect discrepancies between two sources of information, each discrepancy constituting a signal. Each subject performed the task under one of two conditions of signal frequency, high (HSF) and low (LSF). In the HSF condition there was a total of forty-eight signals to be detected during the task and in the LSF condition a total of twenty-four. Each subject performed the task in either noise (95 db) or quiet, in which extraneous noise was masked by a noise level of 70 db.

Subjects

The Maudsley Personality Inventory was administered to eighty-six women undergraduates. From this group twenty-four introverts (mean E scale score 14·37, S.D. 3·73, mean N scale score 28·80, S.D. 10·91) and twenty-four extraverts (mean E scale score 33·50, S.D. 4·11, mean N scale score 22·00, S.D. 11·01) were selected. Each group was further subdivided into four groups of six subjects, making eight experimental groups in all. The eight groups were matched as closely as possible on N scale scores. Subjects were tested between 10 a.m. and noon in groups of twelve.

Apparatus

The task used was a visual cancellation task, consisting of a series of 1800 digits, from 1 to 9 inclusive, selected from a table of random numbers. Consecutive divisions of 360 digits were drawn on five recording traces and fitted on the drum of a spiral kymograph in turn. The guideline for the digits was obtained by attaching the pen of the kymograph to a 1 sec timer with the drum revolving at a speed of 12 mm/sec. One number was hand drawn in the first half of each 1 sec space by means of a template with a window of dimension 10 × 6 mm. The five traces were used consecutively in the experiment by the alternate use of two drums on which the traces were mounted. A short break of 6–7 sec occurred each time the drums were changed over. The display was situated in a small room adjoining the subjects' room and relayed to it by means of closed-circuit television. The camera was focused on a window 10 × 6 mm cut in a piece of white cardboard and attached to the kymograph by means of a clamp. During the experiment the kymograph drum revolved at a speed of 9 mm/sec and each digit was visible on the screen for $\frac{8}{9}$ sec. The digits on the traces were magnified ×9 on the 30 in. screen. The task lasted for 32 min and was continuously watched on a television monitor in the experimental room. In the noise condition white noise was relayed from the television set in the subjects' room into the open field at a level of 95 db. In the quiet condition the level was 70 db.

Procedure

Subjects were instructed to check the order of the digits, as they appeared on the screen, against a typescript of digits and to note any discrepancy between the digits on the screen and those on the script by cancelling the digit on the script and replacing it by the appropriate digit on the screen. Each discrepancy constituted a signal. In the LSF condition there were six signals in each 8 min subperiod of the task and in the HSF condition there were twelve. Alternate HSF signals were omitted in the LSF condition. The temporal arrangement of signals in each subperiod of the task was the same. The total of 1800 digits on the typescript was divided into eight sheets of 225 digits arranged in twenty-five rows of nine. Two 'dummy' sheets were added to these eight, to minimize the occurrence of the 'end-spurt' effect observed when the length and present point in the vigil are known (Bergum & Lehr, 1963). Subjects were also asked to remove their watches. Two practice periods were given after the task had been explained to subjects. In the first, which lasted for 2 min, the occurrences of signals were pointed out to subjects. They were then invited to ask questions. The second practice period lasted for 8 min, and the task followed after a short break. In this second period knowledge of results was not given. Two response measures were taken; correct detections and errors of commission, i.e. when a subject reported the detection of a signal when one had not in fact been presented.

RESULTS

Each of the three variables of the experiment, extraversion or introversion, noise or quiet, high or low signal frequency, was examined in a 2 × 2 × 2 factorial design. The four time-periods of the task were also taken into consideration. The data were treated by means of a trend analysis (Edwards, 1960). The detection scores from the LSF conditions were doubled to facilitate comparison with those from the HSF conditions. The analysis for correct detections is shown in Table 1 and for errors of commission in Table 2. In both analyses, mean squares were tested against the corresponding between-subjects or within-subjects residuals. In the analysis of correct detections the within-subjects residual includes the E × N × S × P interaction which was not calculated since it was clear that most of the variance was accounted for by the other interactions. The residual is therefore larger than the usual subjects pooled × periods error term and the F test more stringent. It is most unlikely that the two non-significant F ratios would have been affected by this. Fig. 1 shows the mean correct detection and commissive error scores of introverts and extraverts plotted against time for the four conditions of the experiment.

Table 1 and Fig. 1 suggest four principal findings for correct detections. (*a*) Significant effects appear only when time at work is taken into account. The significant periods effect indicates that the shape of the overall performance curve changed significantly with time at work. Performance, averaged over the eight experimental conditions, declined with time on task and the linear component of the performance

Table 1. *Trend analysis of correct detection scores for introverts and extraverts over the four conditions of the experiment*

Source of variation	S.S.	D.F.	M.S.	F
	Between subjects			
Extraversion (E)	23·37	1	23·37	3·22
Noise levels (N)	18·13	1	18·13	2·50
Signal frequency (S)	2·75	1	2·75	0·38
E × N	17·75	1	17·75	2·45
E × S	0·89	1	0·89	0·12
N × S	24·63	1	24·63	3·40
E × N × S	10·18	1	10·18	1·40
Residual (subjects pooled)	290·04	40	7·25	—
Total	387·74	47	—	—
	Within subjects			
Periods (P)	19·05	3	6·35	3·29*
E × P	54·78	3	18·26	9·46***
N × P	63·86	3	21·29	11·03***
S × P	16·31	3	5·44	2·81*
E × N × P	43·48	3	14·46	7·49***
E × S × P	5·84	3	1·95	1·01
N × S × P	3·62	3	1·21	0·63
Residual (including E × N × S × P)	237·91	123	1·93	—
Total	444·75	144	—	—

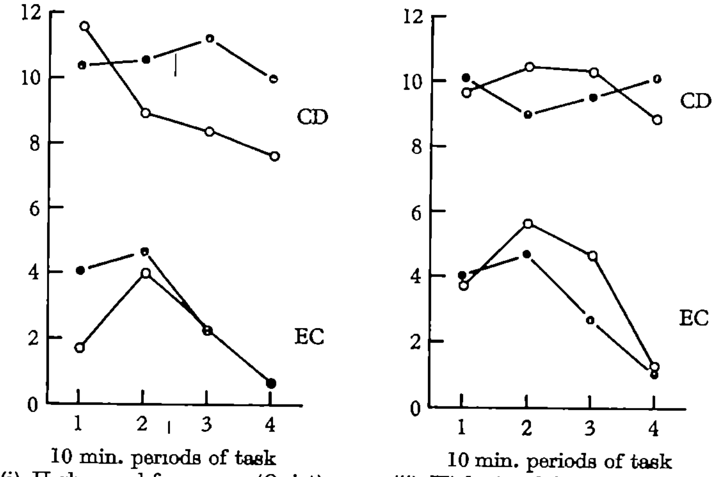
* $P < 0·05$; *** $P < 0·001$.

Table 2. *Trend analysis of errors of commission for introverts and extraverts over the four conditions of the experiment*

Source of variation	S.S.	D.F.	M.S.	F
	Between subjects			
Extraversion (E)	0·02	1	0·02	0·005
Noise levels (N)	1·01	1	1·01	0·28
Signal frequency (S)	6·01	1	6·01	1·67
E × N	15·19	1	15·19	4·23*
E × S	1·04	1	1·04	0·28
N × S	11·04	1	11·04	3·21
E × N × S	37·42	1	37·42	10·42***
Residual (subjects pooled)	143·50	40	3·59	—
Total	215·23	47	—	—
	Within subjects			
Periods (P)	70·85	3	23·62	16·75***
E × P	3·75	3	1·25	0·88
N × P	1·80	3	0·60	0·42
S × P	4·75	3	1·58	1·12
E × N × P	7·74	3	2·58	1·82
E × S × P	26·72	3	8·91	6·24**
N × S × P	22·37	3	7·62	5·40**
E × N × S × P	35·57	3	11·86	8·41***
Residual	169·70	120	1·41	—
Total	343·75	144	—	—

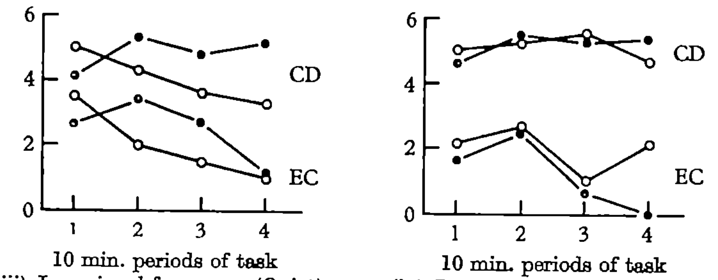
* $P < 0·05$; ** $P < 0·01$; *** $P < 0·001$.

curve is significant ($F = 7{\cdot}76$; D.F. 1, 123; $P < 0{\cdot}025$). The significant extraversion × periods interaction indicates that the effect of time at work on the shape of the performance curve is significantly different for introverts and extraverts, the latter showing a greater tendency to detect fewer signals as the task progresses. (*b*) The significant noise × periods interaction would appear to indicate that noise improves performance to a greater extent towards the end of the task. (*c*) The signal frequency × periods interaction would appear to indicate that in the first, third and fourth periods of the task, HSF groups are superior to LSF groups, whereas in the second period, LSF groups are superior to HSF groups. (*d*) The significant extraversion × noise × periods interaction indicates that the differential shift in efficiency of introverts and extraverts with time at work depends on the noise level. The decrement found in the performance of extraverts at both levels of signal frequency in quiet disappeared in noise, although the performance curves for introverts remained very much the same under both conditions. In other words, extraverts are more likely than introverts to show decrement in quiet, a finding which is similar to that of Bakan, Belton & Toth (1963). This difference is not apparent in noise.



(i) High signal frequency (Quiet)

(ii) High signal frequency (Noise)



(iii) Low signal frequency (Quiet)

(iv) Low signal frequency (Noise)

Fig. 1. Mean number of correct detections (CD) and mean number of errors of commission (EC) made by introverts (filled circles) and extraverts (open circles) in the four conditions of the experiment.

Chi-square analyses of decrement were made, following a procedure adopted by

Bakan *et al.* (1963): decrement was indicated by a subject detecting fewer signals in periods 3 and 4 than in periods 1 and 2. Since no differences were found between comparable high and low signal frequency groups with respect to this measure, the results were pooled and the analyses carried out between the variables of extraversion and noise level only. Only one significant result was obtained: the comparison between introverts (quiet) and extraverts (quiet) yielded a χ^2 value of 8·17 ($P < 0{\cdot}005$). indicating that in quiet more extraverts (10/12) than introverts (2/12) show decrement. This finding is in agreement with that reported by Bakan *et al.* (1963) and supports the significant extraversion × noise × periods interaction noted in the trend analysis.

When the mean level of performance of introverts and extraverts was considered in each of the four task conditions, introverts were found to detect significantly more signals than extraverts in quiet at both low and high signal frequencies (Mann–Whitney U tests, $P < 0{\cdot}05$ and $P < 0{\cdot}03$ respectively). Extraverts improved significantly with noise under LSF conditions (Mann–Whitney U test, $P < 0{\cdot}008$) but their improvement with noise under HSF conditions failed to reach significance. Noise had little or no effect on the performance of introverts at either level of signal frequency, and doubling the signal frequency did not significantly alter the performance level of either group in noise or quiet.

For errors of commission, it can be seen from Table 2 that there are two significant main effects, the extraversion × noise and the extraversion × noise × signal frequency interactions. Introverts made more errors of commission in quiet than in noise, while extraverts made more in noise than in quiet. Under LSF conditions extraverts made the same number of errors of commission in noise as in quiet, while introverts made twice as many in quiet as in noise. Under HSF conditions introverts made about the same number of errors in quiet as in noise while extraverts made nearly twice as many in noise as in quiet.

When time at work is taken into account there is a significant effect of periods, as with correct detections. The number of errors of commission increased from the first period to the second and declined sharply from the second period to the fourth. The linear component of the trend is significant ($F = 35{\cdot}75$; D.F. 1, 120; $P < 0{\cdot}001$). This decline in commissive errors with time on task can perhaps be interpreted as the result of practice (McGrath, 1963); subjects gradually learn what constitutes a signal and what does not and their criterion for responding becomes more cautious.

The other three significant interactions appear to have the following implications. (*a*) Under quiet conditions more errors were made under LSF conditions than under HSF conditions, in each period of the task. This difference was most marked in the first period and gradually diminished during the succeeding three periods. Under noise conditions, the HSF and LSF trend lines fall between the two trend lines for quiet conditions. The difference between the two trend lines in noise is much less than the difference in quiet. (*b*) Extraversion also interacted with signal frequency and time at work. The trend lines for introverts in both signal frequency conditions lie between those for extraverts. Extraverts made more errors under HSF conditions and these were substantially reduced under LSF conditions. Introverts made slightly more errors under LSF conditions than under HSF conditions. Noise also affected this relation, as might be expected from the significant E × N × S interaction.

Summarizing these results, correct detections decreased over time, but this finding can be attributed entirely to the results for extraverts working under quiet conditions. They showed a considerable decline in performance, while extraverts under noise, and introverts under noise or quiet, showed essentially no change with time at work.

As far as total scores for errors of commission are concerned, extraverts made most errors under the condition characterized by noise and HSF, while introverts made most errors under the condition of quiet and LSF. Thus the number of errors of commission was greatest for extraverts in the most 'stimulating' condition and for introverts in the least 'stimulating' condition. When changes over time are taken into account, it is apparent that the number of errors decreased over time but the extent of the decrement varied with different experimental conditions. Signal frequency affected the steepness of the decline in that the decrement was greater for LSF than for HSF groups, but only under quiet conditions or when extraverts alone are considered.

Table 3. *Mean number of responses (correct detections + errors of commission) for each subject*

	Low signal frequency		High signal frequency	
	Quiet	Noise	Quiet	Noise
Introverts	29·33	25·50	49·16	45·16
Extraverts	22·66	29·33	41·50	48·33
	Correct detections as a percentage of all responses			
Introverts	67 %	80 %	86 %	86 %
Extraverts	65 %	73 %	90 %	84 %

Finally, the total number of responses (correct detections + errors of commission) and correct detections expressed as a percentage of all responses for introverts and extraverts under the four task conditions are shown in Table 3. The percentage of all responses that were correct detections made by extraverts in quiet at both levels of signal frequency was about the same as that for introverts. The difference between the number of correct detections made by introverts and extraverts in quiet can be accounted for by the fact that introverts in quiet made more responses; it is not necessary to suppose that they are more likely to be correct when they do respond. In noise the rate of responding increased for extraverts and decreased for introverts by about the same amount at both signal frequency levels. In HSF conditions, although the percentage of responses that were correct detections increased, compared with LSF conditions, there is again very little difference between introverts and extraverts. In noise introverts respond less but are more correct when they do, while extraverts simply respond more.

DISCUSSION

The hypotheses put forward in the introduction to this paper are in the main confirmed and the results would seem best suited to an interpretation in terms of arousal. However, some qualifications are indicated. The effects of noise, signal frequency, extraversion and time at work interact in a complex manner. It is therefore unrealistic to specify what effect each of the variables is exerting upon performance, since any such effect will be highly dependent upon other experimental variables. This finding underlines the difficulty of comparing results from different studies unless they are obtained under identical conditions.

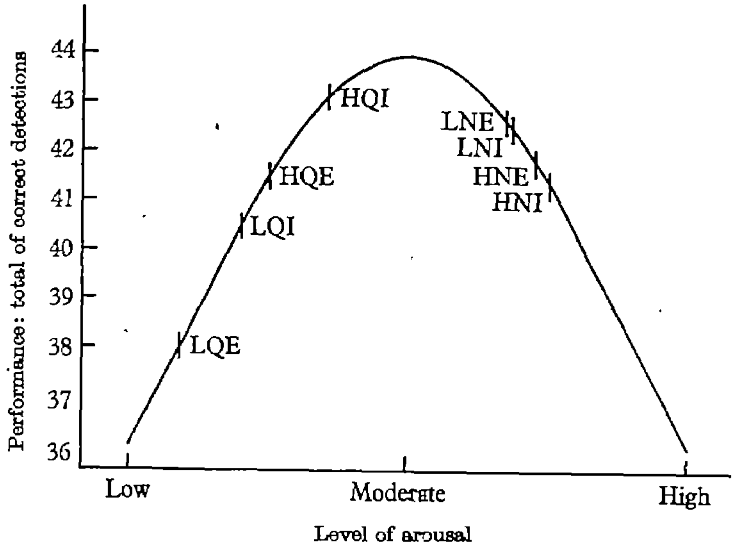


Fig. 2. Total correct detection scores for the eight experimental groups plotted against the hypothesized level of arousal. L, low signal frequency; H, high signal frequency; Q, quiet; N, noise; I, introverts; E, extraverts.

Three assumptions were stated in the introduction: that introverts are more aroused than extraverts, that noise is more arousing than quiet and that a high signal frequency is more arousing than a low one. The relation between arousal and performance was considered to take the form of an inverted U-shaped curve. Given these assumptions, and the further assumption that adding noise is more arousing than increasing the signal frequency, the eight groups can be ordered in terms of arousal as shown from left to right on the abscissa in Fig. 2. When the total number of correct detections is plotted on the ordinate the result can be seen to fit an inverted U. However, no such relation applies to errors of commission or to the total number of responses. Arousal must therefore be envisaged as affecting correct detections in a predictable manner but the way in which it affects the other two response measures, if at all, remains unclear. In any case there is no independent measure of arousal in Fig. 2.

Moreover, it is difficult to explain extraversion–introversion differences in terms of arousal, if this is taken to imply an underlying physiological variable. Reason & Benson (personal communication) using forty-eight subjects, have found no correlation between Maudsley Personality Inventory scores and either (*a*) the resting level of heart rate, muscle tension or skin resistance, or (*b*) the difference between the resting level and the working level during performance of a tracking task. The former variables could be regarded as measures of arousal and the latter as measures of arousability. Thus extraverts and introverts could be held to differ in arousal only if by arousal is meant a cognitive state not directly related to activation as measured physiologically.

In the case of the individual variables in this experiment there are at least equally efficient alternatives to the arousal explanation. However, since the significant effects to be explained are predominantly interactional, an arousal theory has the advantage of accounting for the complex effects of different factors in terms of variation along a single hypothetical continuum. Thus, in spite of some difficulties of interpretation, an arousal theory of vigilance performance provides a reasonably parsimonious account of the present findings, although a conclusive test of its adequacy remains to be made.

The authors are grateful to members of the Department of Psychology, University of Leicester, and to Dr D. W. J. Corcoran, for their comments on earlier drafts of this paper. This research was supported in part by a grant from the Medical Research Council.

REFERENCES

BAKAN, P. (1959). Extraversion–introversion and improvement in an auditory vigilance task. *Brit. J. Psychol.* **50**, 325–32.

BAKAN, P., BELTON, J. A. & TOTH, J. C. (1963). Extraversion–intraversion and decrement in an auditory vigilance task. In *Vigilance: a Symposium*, ed. D. N. Buckner and J. J. McGrath, pp. 22–33. New York: McGraw-Hill.

BERGUM, B. O. & LEHR, D. J. (1963). End spurt in vigilance. *J. exp. Psychol.* **66**, 383–5.

BROADBENT, D. E. (1963). Possibilities and difficulties in the concept of arousal. In *Vigilance: a Symposium*, ed. D. N. Buckner and J. J. McGrath, pp. 184–98. New York: McGraw-Hill.

CORCORAN, D. W. J. (1962*a*). Noise and loss of sleep. *Quart. J. exp. Psychol.* **14**, 178–82.

CORCORAN, D. W. J. (1962*b*). Doubling the rate of signal presentation in a vigilance task during sleep deprivation. *J. appl. Psychol.* **47**, 412–15.

CORCORAN, D. W. J. (1963). Individual differences in performance after loss of sleep. Unpublished Ph.D. thesis, University of Cambridge.

CORCORAN, D. W. J. (1964). Changes in heart rate and performance as a result of loss of sleep. *Brit. J. Psychol.* **55**, 307–14.

CORCORAN, D. W. J. (1965). Personality and the inverted-U relation. *Brit. J. Psychol.* **56**, 267–74.

DUFFY, E. (1957). The psychological significance of the concept of arousal or 'activation'. *Psychol. Rev.* **64**, 265–75.

DUFFY, E. (1962). *Activation and Behavior.* New York: Wiley.

EDWARDS, A. L. (1960). *Experimental Design in Psychological Research.* New York: Holt, Rinehard and Winston.

EYSENCK, H. J. (1959). *The Manual of the Maudsley Personality Inventory.* London: University of London Press.

HEBB, D. O. (1955). Drives and the c.n.s. (conceptual nervous system). *Psychol. Rev.* **62**, 243–54.

HERON, A. (1956). A two-part personality measure for use as a research criterion. *Brit. J. Psychol.* **47**, 243–51.

MALMO, R. B. (1959). Activation: a neuro-psychological dimension. *Psychol. Rev.* **66**, 367–86.

MCGRATH, J. J. (1963). Some problems of definition and criteria in the study of vigilance performance. In *Vigilance: a Symposium*, ed. D. N. Buckner and J. J. McGrath, pp. 227–46. New York: McGraw-Hill.

(*Manuscript received* 27 *September* 1965)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 391–392
Printed in Great Britain

AGE DIFFERENCES IN ERRORS OF COMMISSION

BY G. S. TUNE

Medical Research Council Unit for Research on Occupational Aspects of Ageing, Department of Psychology, University of Liverpool

Subjects listened to a recording of groups of 10 digits (spoken at the rate of 1/sec) for 40 min; each was followed by 10 sec silence. The task was to report whether or not three consecutive and different odd digits occurred and to report the decision in the 10 sec silent periods. Correct detections did not vary either as a function of time or age. The number of errors of commission increased with time on task and the older subjects made significantly more of these errors than did the younger subjects. The results are discussed briefly in terms of the arousal hypothesis.

The absence of reported age differences in correct detection scores in vigilance tasks (Davies & Griew, 1965) would follow from the arousal hypothesis (Broadbent, 1963) and if older people are less aroused than are younger. In order to perform efficiently (achieve a high correct detection score) the older subjects would have to maintain a higher than basal level of arousal. This they would achieve most readily by responding frequently, even to the extent of reporting that a signal had occurred when in fact it had not (making an error of commission). Consistently reporting many signals as 'wanted' when they were not would tend to boost the correct detection score. At the same time, however, a high frequency of responses would produce a high proportion of errors of commission. It is suggested, therefore, that in a vigilance situation older subjects should give more responses than younger subjects, and that an age difference should not be evident in the correct detection score (the usual measure of performance) but in the error of commission score. The present experiment was designed to examine this point of view with reference to the performance of comparable groups of older and younger subjects.

METHOD

Subjects listened to a tape recording of groups of 10 digits (the numbers 1 to 9 inclusive were used in a quasi-random order) spoken at the rate of 1 digit/sec. Each group was followed by 10 sec silence. A 'signal' was the occurrence somewhere in the group of ten (the position was randomized) of three, consecutive and different odd digits (e.g. 573, 195). After they had listened to each group of ten digits, subjects placed a tick in one of two columns headed either 'No' (a wanted combination of numbers did not occur) or 'Yes' (a wanted combination of numbers did occur) on a printed sheet. In the latter case the subjects were encouraged to write the combination they had heard in a third column. It was stressed that this was of little importance and errors in transcribing were not counted against the subject in the scoring procedure. Each subject listened to 120 groups of digits over a period of 40 min. Eight of these (two in each 10 min period) contained a wanted combination. Twenty-eight subjects were used, fourteen older (mean age = 60·6 yr) and fourteen younger (mean age = 38·9 yr).

RESULTS AND DISCUSSION

The distribution of all the 'Yes' responses classified by type, age and time periods of 10 min is shown in Table 1.

A test of trend on the raw correct detection scores showed that neither age, time nor the age by time interaction were significant. Similarly a Mann–Whitney U test on the total correct detection scores for the subjects in each group showed no significant difference. These findings are in accord with previous findings reviewed elsewhere (Davies & Griew, 1965). A test of trend on the raw error of commission scores indicated that age was a significant variable ($F = 6{\cdot}91$; D.F. 1, 26; $P < 0{\cdot}05$) as was time ($F = 20{\cdot}92$; D.F. 3, 79; $P < 0{\cdot}001$). The interaction term was significant at a lower level ($F = 2{\cdot}623$; D.F. 3, 78; $P < 0{\cdot}10$). Mann–Whitney U tests demonstrated that the older subjects made significantly more errors of commission ($P < 0{\cdot}01$) and more total responses ($P < 0{\cdot}01$).

Table 1. *The distribution of 'Yes' responses (in terms of the percentages of the total number of responses made) classified by type, age and time*

10 min time periods	Correct detections		Errors of commission		Totals
	Old	Young	Old	Young	
1	4·1	4·9	16·1	6·5	31·6
2	4·3	4·3	11·5	1·2	21·3
3	4·3	4·4	17·5	3·8	30·0
4	4·4	4·4	8·1	0·2	17·1
Totals	17·1	18·0	53·2	11·7	100·0

These data tend to confirm the suggestions made above and to show that the older subjects achieved parity with younger in terms of the number of correct detections made, mainly because of the large number of errors of commission that they made. Consideration only of the correct detection scores would have concealed the age differences, a fact which emphasizes the usefulness and importance of this score as a measure of vigilance performance.

I should like to thank Mr R. Aldridge-Morris who carried out some of the experimental work.

REFERENCES

BROADBENT, D. E. (1963). Chapter 13 in *Vigilance: A Symposium*, ed. D. N. Buckner and J. J. McGrath. New York: McGraw-Hill.

DAVIES, D. R. & GRIEW, S. (1965). Age and vigilance. In *Ageing, Behaviour and the Nervous System*, ed. A. T. Welford and J. E. Birren. Springfield, Illinois: C. C. Thomas.

(*Manuscript received* 22 *April* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 393–396
Printed in Great Britain

VISUAL MASKING AS A FUNCTION OF PERSONALITY

BY R. J. McLAUGHLIN* AND H. J. EYSENCK
Institute of Psychiatry, University of London

Visual masking was related to the personality dimensions of introversion–extraversion and neuroticism. It was predicted and found that extraverts have a higher masking threshold than introverts. The function of reticulo-cortical arousal and inhibition mechanisms is discussed in relation to extraversion–introversion as a possible source for an explanation of this phenomenon.

Visual masking, meta-contrast, or the suppression of the primary visual stimulus, are terms which have been applied to a phenomenon in which one visual stimulus (S_1) is applied for a brief period (10–30 msec) and followed by another brief visual stimulus (S_2) after a variable inter-stimulus interval (ISI). With ISI of 40 msec or less only S_2 is seen, and S_1 is suppressed or 'masked'; with ISI of 100 msec or more, both stimuli are seen. For most subjects there is an intermediate value of the ISI (masking threshold) at which S_1 ceases to be suppressed, and is seen. This threshold can be established with considerable accuracy (Holland, 1963; Pollack, 1965) and forced-choice methods do not seem to have any advantage over conventional procedures (Heckenmueller & Dember, 1965). Other phenomena of a similar kind, such as the Bidwell effect (Eysenck & Aiba, 1957; Aiba, 1963) appear to obey similar laws, and visually evoked potentials have been shown to correlate with the psychological events of visual masking (Donchin & Lindsley, 1965).

Since S_2 appears to ***inhibit*** the perception of S_1, it has been argued (Eysenck, 1957) that personality differences in inhibitory potential should be relevant to this phenomenon, as should c.n.s. stimulant and depressant drugs. In terms of physiological arousal theory, introverts are postulated to be in a state of higher cortical arousal than extraverts (Eysenck, 1963, 1964) and should thus be able to maintain the discrimination between S_1 and S_2 more efficiently than extraverts. Similarly, stimulant drugs, by increasing cortical arousal, should increase discrimination, while depressant drugs, by decreasing arousal, should have the opposite effect. It was predicted therefore that introverts (and persons given stimulant drugs) would have lower masking thresholds than extraverts (and persons given depressant drugs). The drug prediction has been verified several times (Eysenck, 1957; Aiba, 1963; Holland, 1963): the purpose of the present investigation was to test the personality prediction.

METHOD

Subjects

Sixty-four paid volunteers were tested on the Eysenck Personality Inventory (EPI) (Eysenck & Eysenck, 1964) and put into four groups: neurotic extraverts (NE); stable extraverts (SE); neurotic introverts (NI) and stable introverts. Eight subjects were eliminated from further testing because they obtained scores from 11 to 13 on the E-I scale (mean = 12·0) and were thus not easily classifiable as either introverted or extraverted. One other subject was eliminated at this point because his 'lie scale' score was too high (L = 6) for his personality score to be accepted as valid.

---

* Now at Saint Louis University, Saint Louis, Missouri.

Seven subjects were eliminated during the experiment, five for having an error rate above 10 %, indicating that they were reporting the test stimulus when it was not being presented, and a further two subjects for not reporting the presence of the test stimulus below an ISI of 130 msec, which might be an indication that the subject did not have normal visual acuity.

The remaining forty-eight subjects, eleven females and thirty-seven males, were equally divided, twelve in each of the four personality groups. The mean age of subjects was 25·71 years (s.d. = 8·29) and all reported that they had normal or corrected-to-normal visual acuity.

Apparatus

The instrument, previously described by Holland (1963, pp. 72–5), was an electronic Dodge-type tachistoscope with variable exposure and interval between exposures. The exposed fields were illuminated at 0·155 log. *L* by four 9 in. 6 W fluorescent tubes arranged around a rectangle and shaded to illuminate a 7 in. square. The stimuli presented were a 20 mm black disk (visual angle 0·46°) and a black annulus with inner and outer diameters of 20 mm and 58 mm (visual angle 1·26°) both presented on a white ground. The disk was viewed binocularly through, and the annulus reflected from, a half-silvered mirror placed at 45° to the line of regard. The sequence of stimulus presentations was: (1) test figure 15 msec; (2) ISI, variable; (3) masking figure, 15 msec. The darkened fields between the trials contained a pin-point source of red light superimposed on the centre where the disk and annulus appeared. This point served as a focusing stimulus when the subject was told that the lights were to be flashed.

The tachistoscope was modified to carry two display wheels which could easily be rotated into position and held by a spring-loaded ball-bearing and dimple giving very accurate positioning. Hence it was easy to select between a present or absent test stimulus and a present or absent masking stimulus within a matter of 2 or 3 sec.

Procedure

Each subject was tested individually. The procedure followed that reported by Holland (1963) with sometimes two exceptions. First, a measurement of the subject's responses was made on trials in which the annulus was shown but the disk was not (to establish error rate). Secondly, subjects were required to rate on a three-point scale (1 = 'positive'; 2 = 'fairly certain'; 3 = 'not sure') their confidence in having seen the disk on those trials in which they reported its appearance. An ascending method of limits was used proceeding from an ISI of 40 msec and increasing in steps of 5 msec until the subject responded correctly on two consecutive trials that he was positive (saying '1') that he had seen the test stimulus. It was necessary to limit the range of ISI tested and only to test in a series of ascending limits to maximize the reliability of the measure, because of the limited availabliity of subjects and the problems of fatigue that would be expected if the session exceeded 45 min. After five consecutive threshold measurements the subject was given a rest period of approximately 5 min before a final series of five further threshold measurements was made.

Results

The masking threshold was computed for each subject by averaging the ISI of the first response, for each of the ten trials, for which the subject correctly reported that he was positive the disk appeared. Table 1 summarizes the masking thresholds for the four personality groups. A 2 × 2 analysis of variance was applied to the thresholds,

Table 1. *Mean masking thresholds* (*and* s.d.) *for each of the four personality groups* (n = 12); *for all four groups combined, mean threshold* = 71·11 (s.d. = 12·26, n = 48)

	Stable		Neurotic	
	Mean	s.d.	Mean	s.d.
Extraverts	74·33	12·73	75·04	13·19
Introverts	68·83	10·68	66·25	11·50

comparing extraversion–introversion with high and low levels of neuroticism. The comparison of introversion with extraversion yielded the only significant effect ($F = 4{\cdot}21$; D.F. 1, 44; $P < 0{\cdot}05$) with extraverts having a higher threshold than introverts. Bartlett's test showed no significant heterogeneity of variance.

The correlation between the threshold for the first five trials and the second five was 0·96 for all subjects, showing high intra-subject reliability of the masking threshold measurements.

DISCUSSION

A possible explanation of masking phenomena may be found in Granit's (1955) concept of pre-excitatory inhibition (Eysenck, 1957); it seems likely that higher centres play an important part in this activity (Aiba, 1963). Donchin & Lindsley (1965), who have studied averaged evoked cortical potentials to pairs of flash stimuli and who found that physiological masking occurred with short ISI, have raised the question of the localization of the stage in the visual system at which the neural effect produced by S_2 interferes with the effect elicited by S_1; they consider four possibilities: (1) retina (receptor cell layer to ganglion cell layer), (2) lateral geniculate body (relay and other functions), (3) cortical structures (primary and secondary), and (4) non-specific sensory system (reticular and thalamic). Eysenck (1963, 1964) has argued for the relevance of the reticular system in connexion with personality differences along the extraversion–introversion axis, and data from experiments with drugs are in good agreement with this hypothesis (Killam, 1962).

The work of Shagass & Schwartz (1963*a*, *b*) appears to be relevant. They administered pairs of stimuli and recorded the evoked potentials, thus making it possible to plot the cycle of recovery of responsiveness. They found that the early portion of the recovery cycle for the primary somatosensory response tended to be biphasic in form. The first phase of recovery occurs very early, the amplitude of the second response equalling that of the first, or exceeding it before 20 msec. There is then a phase of diminished responsiveness, followed by a longer phase of full recovery, usually reaching a peak at about 100–130 msec. Dysthymics (introverted neurotics) were found to have higher recovery ratios than psychopaths (personality disorders—neurotic extraverts), i.e. they recovered more quickly. These date are in good agreement with those reported here, and suggest that the processes involved are essentially similar.

Another phenomenon which appears closely relevant is critical flicker fusion (Eysenck, 1957). There is evidence that introverts have higher C.F.F. thresholds (i.e. discriminate better between flashes at high frequencies) than extraverts, and that stimulant drugs improve discrimination, while depressant drugs have an opposite effect (Eysenck, 1967). Aiba (1963) has compared C.F.F. with masking and has drawn attention to some of the similarities, and to the possibilities of accounting for both along similar theoretical lines. It cannot be said that any particular theory in this field is at present clearly indicated as more correct than any of the others; indeed, it seems likely that these phenomena represent the confluence of neural impulses originating from several different points. Nevertheless, there is increasing evidence (*a*) that reticulo–cortico arousal and inhibition mechanisms play a part in the causation of the phenomena themselves, and (*b*) that personality differences along the extraversion–introversion axis are meaningfully related to masking and to C.F.F. effects through the mediation of the reticular system.

We are indebted to the Maudsley and Bethlem Royal Research Fund for financial assistance.

REFERENCES

AIBA, S. (1963). The suppression of the primary visual stimulus. In *Experiments with Drugs*, ed. H. J. Eysenck. Oxford: Pergamon Press.

DONCHIN, E. & LINDSLEY, D. B. (1965). Visually evoked response correlates of perceptual masking and enhancement. *Electroenceph. clin. Neurophysiol.* **19**, 325–5.

EYSENCK, H. J. (1957). *Dynamics of Anxiety and Hysteria*. London: Routledge and Kegan Paul.

EYSENCK, H. J. & AIBA, S. (1957). Drugs and personality. V. The effects of stimulant and depressant drugs on the suppression of the primary visual stimulus. *J. ment. Sci.* **103**, 661–65.

EYSENCK, H. J. (1963). The biological basis of personality. *Nature, Lond.* **199**, 1031–4.

EYSENCK, H. J. (1964). Biological factors in neurosis and crime. *Scientia* **58**, 1–11.

EYSENCK H. J. (1967). *The Biological Basis of Personality*. Springfield: C. C. Thomas.

EYSENCK, H. J. & EYSENCK, S. B. G. (1964). *Manual of the Eysenck Personality Inventory*. London: University of London Press.

GRANIT, R. (1955). *Receptors and Sensory Perception*. New Haven: Yale University Press.

HECKENHUELLER, E. G. & DEMBER, W. N. (1965). A forced-choice indicator for use with Werner's ring pattern in studies of backward masking. *Psychonom. Sci.* **3**, 167–8.

HOLLAND, H. C. (1963). 'Visual masking' and the effect of stimulant and depressant drugs. In *Experiments with Drugs*, ed. H. J. Eysenck. Oxford: Pergamon Press.

KILLAM, E. K. (1962). Drug action on the brain stem reticular formation. *Pharmacol. Rev.* **14**, 175–224.

POLLACK, R. H. (1965). Backward figural masking as a function of chronological age and intelligence. *Psychonom. Sci.* **3**, 65–66.

SHAGASS, C. & SCHWARTZ, M. (1963*a*). Cerebral responsiveness in psychiatric patients. *A.M.A. Arch. gen. Psychiat.* **8**, 177–89.

SHAGASS, C. & SCHWARTZ, M. (1963*b*). Neurophysiological disfunctions associated with psychiatric disorders. *Psychiat. Res. Rep.* **17**, 130–52.

(*Manuscript received* 22 *April* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 397–403
Printed in Great Britain

A PSYCHOLINGUISTIC INVESTIGATION OF A RULE THEORY OF 'SENSE'

BY HENRY A. ALKER
Educational Testing Service, Princeton, N.J.

Sommers, formalizing some procedures of linguistic analysis used by Wittgenstein and Ryle, has proposed that the senses of a set of terms may be represented uniquely and unequivocally by a topological tree. This claim was tested by examining empirically its consequence that sets of terms not fitting on a tree must contain some equivocation. Each of four different measures of equivocation supported this prediction. Implications for current research on meaning are discussed.

Recent investigations have determined that linguistic rules influence both the intelligibility and the retention of sentences (Miller & Isard, 1963; Marks & Miller, 1964). Three classes of linguistically organized sets of words were used: (*a*) ungrammatical strings of words that were semantically anomalous, (*b*) grammatically well-formed sentences that were still semantically anomalous, and (*c*) grammatically and semantically well-formed sentences. The differential effect of each of these three classes on auditory intelligibility and retention provided experimental justification for distinguishing between syntactic and semantic rules. This paper offers evidence in support of a particular theory concerning the nature and interrelations of semantic rules.

The focus of the theory tested is a particular violation of a semantic rule known as a 'category mistake' (Ryle, 1949). Sentences containing these mistakes are members of the second class distinguished above, namely, grammatically well-formed sentences that are semantically anomalous. The 'mistake' at issue occurs when the predicate of a sentence assigns to the subject of the sentence some action or property that makes no sense. Examples of this kind of grammatically well-formed sentence are: 'The number 7 is on the table', 'The sound of music is in my soup', and, in the case of a compound predicate, 'She went home in a Cadillac and a flood of tears.' Part of the meaning of a term is exhibited when its permissible and inappropriate category or sense relations are stated.

Sommers (1959) has derived from the writings of Ryle (1949) and Wittgenstein (1953) an explicit model for representing the sense relations of different terms. The model has the topological properties of a tree. In effect Sommers has argued that the sense relations of different terms can be represented as a tree, rather than, for example, by a net or web, *only if* the sense of the term in its various category or sense relations is unequivocal. The topological property of a tree capitalized upon by this model is that there is only one path (without retracing) on the tree between the location of any two senses. In other words, each individual sense, be it of the same or a different term, is uniquely located with respect to its category relations on the tree. Surely the capacity of a model of semantic space to guarantee a distinct location for each difference in sense is no mean achievement. The unique location of terms mapped with respect to their category relations may be illustrated by introducing the notion that a particular sense of a term can be 'higher than' other senses of the same or some other term. Following Sommers (1959), this relation is defined by the presence of a meaningful category relation between the 'higher' term and each of two other terms; whereas, these other terms themselves do not occur together without committing a category mistake. An example of the location of a sense being 'higher than' the location of two other senses is provided in Fig. 1. In that example the term 'explanation' is higher than the sense of 'nuclear fission' and of 'contained in three volumes'. It is permissible to say that 'The explanation is nuclear fission' and 'The explanation is contained in three volumes', but it would be a category mistake to assert in some ordinary (i.e. non-metaphorical or non-facetious) context that 'Nuclear fission is contained in three volumes.'

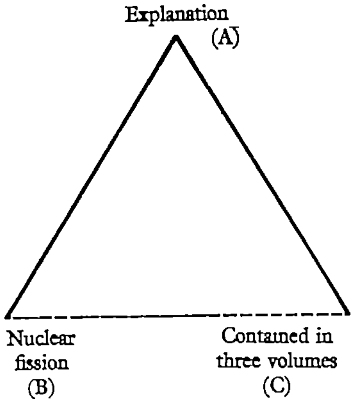


Fig. 1

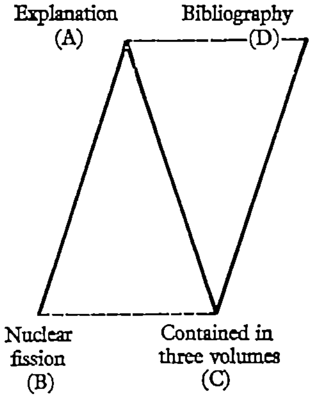


Fig. 2

Fig. 1. A triad with term A higher than terms B and C. The solid lines represent permissible sense relations, the dotted line represents a category mistake.

Fig. 2. The fourth term 'bibliography' exhibits sense relations in violation of the equivocation rule. Term A is higher than C (and B), but term C is higher than A (and D).

The tree model requires that when the sense of a term such as 'explanation' is higher than the two other senses mentioned, neither of these senses can be higher than the sense of 'explanation' with respect to the sense of any fourth term. For if we can find a term that is permissibly sense-related to either 'nuclear fission' or 'contained in three volumes' and some additional term then 'nuclear fission' or 'contained in three volumes' will be, by definition, higher than 'explanation'. 'Explanation', however, was originally located higher than these other terms. When the relation 'higher than' becomes symmetrical in this way the unique path (without retracing) to each sense location is destroyed. Fig. 2 illustrates a putative violation of the tree model or, as it is alternatively called, the equivocation rule. The fourth term, 'bibliography', can easily have as a predicate 'contained in three volumes' but it does not appear literally meaningful to say that 'The explanation is a bibliography'. This U relation of 'bibliography' makes the sense of 'contained in three volumes' higher than the sense of 'explanation' when this latter term is also higher than 'contained in three volumes'. A possible change in sense, which the theory predicts must occur in this case, might reside in the use of the term 'explanation'. The sense in which we speak of the explanation being nuclear fission may be quite distinct from the sense in which an explanation is said to be contained in three volumes. In the former context the phenomenon referred to by the explanation is at issue; in the latter context one is speaking of the presentation of the explanation. It is this condition of the theory, that a change in sense or equivocation must occur when mapping on a tree is not possible, that allows the theory to be tested empirically by obtaining instances where equivocation is expected on theoretical grounds and assessing the equivocation independently.

METHOD

Procedure

The first task for the subjects was to judge the category relations of sixteen words. A sentence frame was provided containing a set of permissible articles, a space for a subject term, a set of permissible forms of the verb 'to be' (past, present, future), another set of permissible articles (identical to the first set), and a space for a predicate complement. The subjects were requested to judge whether or not a set of two words fit into the frame, so as to produce a literally meaningful sentence regardless of whether or not the sentence was likely to be true. The subjects were told that the order in which the words were provided could be reversed, if necessary, in their attempt to construct a meaningful, grammatical sentence. In the event that the two words provided were both adjectives, special instructions were provided. Subjects were to construe the two words as part of a compound predicate complement. In accordance with Sommers' stipulation on this point (Sommers, 1959), the subjects were asked if they could find any word that would fit into the place of the subject in the sentence frame to produce a meaningful sentence with the compound predicate complement. An affirmative finding indicates a permissible relation between the two adjectives. A negative finding determines that the adjectives are incompatible categories.

Materials

The 120 pairs, representing all possible combinations of sixteen words grouped two at a time, were randomized both with respect to the order of the pairs in which they appeared and with respect to the order within each two-membered pair. Following each pair of words were the terms 'meaningful' and 'meaningless' which the subjects circled to indicate their category judgements. The sixteen words used were selected so that each possible combination, using a dichotomous classification, of the following stimulus properties, was represented by one word (adjective or noun), Semantic Differential polarization score (Jenkins, 1960), Thorndike-Lorge, (1944) frequency, and number of dictionary senses (Webster's, 1961). Obtained results are thus not confounded by covariation among these variables. The twenty-eight pairs containing two adjectives were marked with an asterisk to encourage the subjects to consult again, if necessary, the procedure indicated in the instruction sheet concerning this special case. This indication also allowed the subjects to reduce ambiguities arising concerning part of speech in the cases where an adjective might be employed in a secondary usage as a noun.

Subjects

The subjects were forty undergraduates in introductory psychology classes at the University of California, Berkeley. They received credit towards course completion for participating in the experiment.

Measures of change in sense

In order to test the hypothesis that unequivocal senses are mapped by a tree structure of category relations, four different measures of the probability that a word was used in different senses were obtained.

(1) Upon completing their judgements concerning the category relations of the sixteen words, subjects were presented with a form instructing them to rank order the sixteen words with respect

to the probability that the word had been used in different senses while the subject was making his category judgements.

(2) Goldberg (1963) has suggested, in connexion with his discussion of the characteristics of items in a psychological test, that the intersubject variability in responses to an item reflects the equivocality of the item. In this study it is possible to compute an equivocality score by noting the intersubject variability present when subjects respond to a pair of words by judging whether or not a category mistake has been committed. This score has a range for each pair of 0 (complete agreement among the forty subjects about the category relation) to 20 (half the subjects judge that a category mistake is present, half the subjects disagree). It is not possible, of course, to infer which of the two terms in a given pair is the primary contributor to the intersubject variability in judgement. The mean equivocality score from the fifteen different pairs in which a given word occurs, however, is a more suitable measure. Each word enters into pairwise combination with the same words as does each other word, with the negligible difference that no word enters into combination with itself. Different mean equivocality scores consequently reflect differences in intersubject variability with the set of other members in the pair held constant.

(3) A second indirect measure, also based on the notion of intersubject variability being an indicator of equivocality, was obtained from the semantic atlas norms (Jenkins, 1960). This measure, the concept's polarization score, reflects, when low, an absence of intersubject agreement concerning the rating of a concept on the twenty Semantic Differential seven-point scales. A high polarization score reflects consensus about the 'connotative' meaning of the concept. Low polarization scores occasionally also reflect such a consensus concerning the intermediate placement of a term on the seven-point scales. Jenkins (1960) suggests this possibility for 'Lemur'. The use of a battery of equivocation measures should help overcome difficulties of interpretation posed by this confounding.

(4) The final measure used was the number of dictionary senses a term had within its use as one part of speech. Archaic, obsolete and rare senses were not counted. In order to maximize variation between terms, the number of lettered subheadings in the dictionary entry for a term (Webster's, 1961) was counted rather than simply the number of different senses listed. In cases where a term was listed in the dictionary as both an adjective and a noun, this fact was scored as one extra sense for the term in addition to the number of senses the term had in its selected part of speech. When the selected part of speech produced an adjective-adjective pairing, an asterisk reminded the subject of the special instructions applying in this case. Only a small positive relation was expected between this measure and the criterion as the score reflects merely the possibility of the subject having within his semantic repertoire different literal senses for a given term.

RESULTS

Table 1 presents the equivocation scores for the sixteen terms used in this study. These scores were computed from the 227 different obtained U triads exhibiting a 'higher than' relation for which the theory tested in this study made an explicit prediction. Each time the interrelations of a fourth term with the terms of a 'higher than' triad were inconsistent with a tree mapping, all four terms received one point each. All thirteen terms not involved in the triad were inspected for each different triad. For this criterion equivocation score, particular category relations were regarded as incompatible if a majority of the subjects thought the particular pair did not make a literally meaningful sentence in the sentence frame provided. Category relations considered literally meaningful by a majority of the subjects were considered to be permissible. Of the 120 pairs provided, forty-three were defined as incompatible by this criterion, seventy-six were defined as permissible, and one pair, namely 'socialism, sunlight', was indeterminate.

In Table 2 are presented the correlations of the four different predictors of change in sense with the rule theory's deduction concerning this variable. The intercorrelations between these predictors are also presented. The consistency of confirmation among all four predictors is perhaps more impressive than the magnitude of the correlation any one of them attained.

A peculiar improvement occurs if the data involved in the direct measure of change in sense are grouped differently. If it is assumed that the mean rank of rated change in sense of a given term is not maximally discriminating and the original sixteen terms are regrouped into eight ordered groups of two terms each, whose equivocation criterion score is the average of the individual scores of the terms, then the correlation with the criterion scores of the predictor variables, also regrouped into eight ordered groups, increases to +0·71 from +0·43. No appreciable change occurs if the original ranks in each individual protocol are changed from 1 through 16 to 1 through 8. This finding provokes the speculation that some upper limit on the subjects' capacity for processing information is being reached (cf. Miller, 1956), but further investigation is needed to confirm that the additional 32 % of predicted variance is not merely fortuitous.

Table 1. *Equivocation scores: the scores were calculated by giving the A, B, C and D term each one point every time they appeared in a collection of six category judgements violating the equivocation rule* (*see note below*)

Continuous	241	Socialism	133
Bottom	210	Criminal	126
Low	204	Lemur	125
Negro	168	Sunlight	124
America	155	Devil	109
Lofty	152	Rage	101
Lovable	146	Winter	95
Brave	139	Happy	85

The specific patterns of category relations between a fourth term, D, and the three terms A, B, and C in which A is higher than B and C, which indicate equivocation, are as follows. Let $U(AD)$, $U(BD)$, etc., stand for compatible category relations between the two terms in the parentheses and $N(AD)$, $N(BD)$ stand for incompatible category relations between the two terms in the parentheses. Then equivocation 'points' are assigned to A, B, C and D only if (*a*) $N(AD)$ $U(BD)$, (*b*) $N(AD)$ $U(CD)$, or (*c*) $U(AD)$ $N(BD)$ $N(CD)$.

Table 2. *Product-moment correlations of predictors with equivocation score* (*ES*) *and of predictors with each other*

(RE, rated equivocality; IV, intersubject variability; D_4, polarization score; DS, dictionary senses. The positive signs of the correlations with the equivocation score (ES) reflect relations in the predicted direction. The signs of the intercorrelations have been reflected to be consistent with this practice.)

	ES	RE	IV	D_4	DS
ES	—	0·43**	0·30*	0·48**	0·27*
RE	—	—	0·43**	0·04	0·33*
IV	—	—	—	0·05	0·20
D_4	—	—	—	—	−0·01
DS	—	—	—	—	—

$^{*}P < 0·05$, one-tailed test. $^{**}P < 0·01$, one-tailed test.

While the part of speech is related to the equivocation rule criterion, Lorge frequency is not. The point biserial and biserial correlation are, respectively, 0·28 (P < 0·05) and 0·05 (N.S.). The relation of part of speech found here was due to adjectives being more equivocal. Perhaps the subjects found the special instructions to supply subject terms for the compound predicate an additional stimulus to changing the sense of the predicate terms.

As one might expect, several students volunteered after the testing session that they had tended to make metaphorical rather than the literal judgements requested concerning sense relations. It should be clear that, for the theory being tested, this particular deviation from the instructions contributes to the variation in constancy of sense which was the criterion to be predicted.

Discussion

The rule theory supported here has dealt with only a restricted class of grammatically well-formed but semantically anomalous sentences. Sentences that Marks & Miller (1964) have investigated, e.g. 'Colourless green ideas sleep furiously', have combined several category mistakes—N(colourless, ideas), N(green, ideas), N(ideas, sleep) and N(sleep, furiously)—and they have sometimes included implicit contradictions, like the conjunction of 'colourless' and 'green', Consequently, the theory tested here must be viewed as a theory of linguistic competence rather than linguistic performance. The theory provides rules whereby equivocation may be detected in a manner analogous to the way Chompsky's syntactic theory provides for the disambiguation of sentences such as 'I had the books stolen from the library' (cf. Chompsky, 1957). This focus on linguistic competence, a necessary consequence of any rule theory, does not preclude the theory from generating predictions concerning actual linguistic performance.

It can be argued that linguistic behaviour in a context where there is no manifest correctness of the response required still exhibits regularities of a rule-following sort. Concept scale interaction on the semantic differential, for example, is explained by Osgood (1962) in terms of a 'denotative restriction' imposed by the concept on a given scale. To use one of Osgood's examples, the concepts Dawn and United Nations interact with polar adjective scales such as *sober–drunk* and *mature–youthful*. Dawn is rated *sober* rather than *drunk* and *youthful* rather than *mature*. The concept United Nations, on the other hand, while being *sober* rather than *drunk* is *mature* rather than *youthful*. This finding can be accounted for, at least in part, by the hypothesis that subjects making semantic differential ratings are tacitly following the category rules of their language. (Another factor contributing to subjects' responses in the relatively unstructured semantic differential situation is probably their judging the probability of co-occurrence of the referents of the two concepts, cf. Flavell, 1961.) To say that Dawn is *mature* is not literally meaningful. Dawn does not take on responsibilities, pass into senility, qualify for the vote, and so on. The United Nations, on the other hand, in an ordinary sense is *mature*. It has weathered numerous challenges, can take on responsibilities and so on. Tanaka, Oyama & Osgood (1963) explained the concept scale interaction in terms of interaction with concept-class *and* interaction with subject-group, and this explanation points to the value of specifying category relations in attempts at mechanical translation of languages between subject groups.

Summary and conclusions

A rule theory of meaning allows the unequivocal sense of a term to be represented uniquely. The sense is specified by the category relations of the term. In this study the theory was tested by showing that terms whose category relations did *not* fit the specified model were equivocal. It is not claimed that the sense of a term represents its meaning exhaustively. Concepts such as 'blue', and 'not blue' for example, may have identical category relations but clearly different meaning. Their denotation or reference is different. Given the current usage of the term 'connotation' in psychology, it is probably helpful to reserve the term 'sense' for the aspect of meaning treated here. Because sense is a necessary condition for linguistic meaning, it is asserted, however, that two terms with different category relations cannot have the same meaning.

The domain over which the theory claims to apply is that of ordinary, natural language. It is hard to see, however, how a principle that apparently regulates language in use could be irrelevant to the way languages, particularly the use of sentences, is learned. A primitive notion in both contexts would appear to be predication not association.

An early version of this paper was read at the meetings of the Western Psychological Association, Portland, Oregon, U.S.A., on 16 April 1964.

REFERENCES

CHOMPSKY, N. (1957). *Syntactic Structures*. The Hague: Mouton.

FLAVELL, J. H. (1961). Meaning and meaning similarity. II. The semantic differential and co-occurrence as predictors of judged similarity in meaning. *J. gen. Psychol.* 64, 321–35.

GOLDBERG, L. (1963). A model of item ambiguity in personality assessment. *Educ. psychol. Measmt.* 23, 467–92.

JENKINS, J. (1960). Degree of polarization and scores on the principal factors for concepts in the semantic atlas study. *Amer. J. Psychol.* 73, 274–9.

MARKS, L. E., & MILLER, G. A. (1964). The role of semantic and syntactic constraints in the memorization of English sentences. *J. verb. Learn. verb. Behav.* 3, 1–5.

MILLER, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97.

MILLER, G. A., & ISARD S. (1963). Some perceptual consequences of linguistic rules. *J. verb. Learn. verb. Behav.* 2, 217–28.

OSGOOD, C. E. (1962). Studies on the generality of affective meaning systems. *Amer. Psychologist* 17, 1–9.

RYLE, G. (1949). *The Concept of Mind*. London: Hutchinson.

SOMMERS, F. (1959). The ordinary language tree. *Mind* 68, 160–85.

TANAKA, Y., OYAMA, T. & OSGOOD, C. E. (1963). A cross-culture and cross-concept study of the generality of semantic spaces. *J. verb. Learn. verb. Behav.*, 2, 392–405.

THORNDIKE, E. L. & LORGE, I. (1944). *The Teachers Word Book of 30,000 Words*. New York: Teachers College, Columbia University.

Webster's Third New International Dictionary (1961). Springfield, Mass.: Merriam.

WITTGENSTEIN, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.

(*Manuscript received* 30 *December* 1965)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 405–411
Printed in Gerat Britain

DECODING A DECEPTIVE INSTRUCTION

By SHEILA JONES
Medical Research Council, External Staff,
Department of Psychology, University College, London

Equivalent forms of instruction for a task requiring response to one of two equal classes of items were investigated. The class of items requiring response was defined positively (by inclusion) in one form of instruction and negatively (by exclusion) in the other. It was hypothesized that preference for positive rather than negatively defined classes would lead subjects to transform (decode) the exclusion form of instruction into the equivalent inclusion form. Performance of subjects first given the 'exclusion' instruction and afterwards transferred to the 'inclusion' form was compared with a control group, using the latter form throughout. Less than half the subjects decoded the 'exclusion' instruction, and the remainder performed the task at a significantly slower speed both before and after transfer. The error pattern of these subjects indicated that an instruction of the form 'Respond to all items except x and y' generates a spurious set to make a response to the items x and y. The implicit nature of the negative in the qualifier 'except' appeared to be a contributing factor to the deceptive effect of this type of instruction.

The use of a negative qualifier—'except'—in an instruction, for which an equivalent positive form is available, has been shown to reduce efficiency of performance under certain conditions (Jones, 1966). A classification task, in which responses were required to five out of every eight items, was performed at a significantly slower speed under an instruction of the form 'Respond to all except three items' than under the form 'Respond to five items'. For a large domain of items, definition in terms of the smaller class may be unavoidable from the point of view of memory load, or in the interests of verbal economy. Nevertheless, in cases where the size of both classes of items is within the immediate memory span, recourse is frequently made to the exclusion form of definition whenever one class exceeds the other in number of items, e.g. 'The library is open every weekday except Wednesday.'

The advantages of reduced memory load or of verbal economy appear to override considerations of whether the specified class has to be defined positively or negatively. This finding seems to be in conflict with evidence from numerous experiments (Bruner, Goodnow & Austin, 1956; Goodnow & Pettigrew, 1956; Wason, 1960) which show an apparent aversion to the use of negative instances and a preference for positively derived inferences. In particular, Wallach (1959) has shown that subjects prefer to place events in a class defined by inclusion rather than exclusion, in cases where ambiguity exists in the range of acceptance limits between the classes.

The finding that reduced memory load does not always compensate for the inefficiency associated with the use of negative classes raises the question of whether an implicit negative, such as 'except', is regarded as a potential hazard by its users. The subjects in the previous experiment (Jones, 1966) had no knowledge of the total range of items to be encountered and thus had no option but to carry out the task on the basis of the exclusion class. For a defined domain of items, however, the possibility exists of transforming a given instruction, expressed in terms of an exclusion class, to its equivalent positive form.

Preliminary experiments, using psychology students as subjects, in which responses to six out of every ten items were required, showed that about half of them transformed an instruction of the form 'Respond to all except four items' to the equivalent form 'Respond to six items'. The remaining subjects either failed to consider the possibility of such a transformation (or 'decoding') of the instruction, or having considered it, rejected it on the grounds of increased memory load. The introspective reports indicated initial vacillation by some subjects about whether to decode or not. This was partly due to awareness of the difficulties of handling negative statements in previous experiments.

The present experiment was devised to remove the factor of memory load as a variable determining the decision to decode an instruction defined by exclusion. With equal numbers of items in each class, a decision to decode such an instruction should indicate, in the case of naïve subjects, their preference for using inclusion rather than exclusion classes. The decoding behaviour of such subjects can be examined by transferring them from performance of a task under an exclusion instruction to the equivalent inclusion form, and comparing their performance with a control group using the inclusion instruction throughout.

METHOD

Design

Four sets of instructions, two defined in terms of the inclusion class and two in terms of the exclusion class, were used in a task which required four out of every eight digits to be cancelled. In all cases, the instructions were preceded by the statement: 'Only the following digits will appear: 1, 2, 3, 4, 5, 6, 7, 8.' The four sets of instructions were: inclusion instructions: (A) 'Mark the numbers 3, 4, 7, 8', or (B) 'Mark the numbers 1, 2, 5, 6'; exclusion instructions: (A) 'Mark all the numbers except 3, 4, 7, 8', or (B) 'Mark all the numbers except 1, 2, 5, 6'.

Two (experimental) groups were given exclusion instruction (A) or (B), and performed three trials of cancelling digits, after which they were given the equivalent inclusion instruction, (B) or (A), and performed three further trials. Two (control) groups were given inclusion instruction (A) or (B); after three trials the same instruction was repeated and three further trials were performed. Thus the actual responses required of the experimental and the corresponding control groups were the same throughout both sets of three trials.

Materials

Each trial consisted of a page of 144 digits (twelve rows of twelve equally spaced digits). The digits 1–8 occurred with equal frequency and were arranged in a different random order for each of the six trials.

Procedure

The subjects were assigned alphabetically to each form of instruction, in the proportion of two subjects to exclusion instructions to one subject assigned to inclusion instructions. Each subject was tested individually and was given a preliminary demonstration of the method of cancelling to be used (a stroke through each digit). Subjects were asked to work as quickly as possible, consistently with accuracy. They were instructed not to lose time in correcting errors if they occurred.

After the preliminary demonstration and general instructions about the mode of performance, each subject was given a piece of paper with his specific instruction, inclusion (A) or (B) or exclusion (A) or (B), preceded by the information that the range of digits was from 1 to 8. At this stage the experimenter told each subject that the instruction was about to be removed and asked him 'Tell me what you are going to do?'; the subject's response was noted. The instruction was then withdrawn and the subject was presented with the first of three trials. A rest pause of 30 sec, timed with a stop-watch, was allowed between each trial. After three trials a second instruction sheet was given to the subject and withdrawn after 1 min. The subject then performed a further three trials as before.

Subjects

The subjects were fifty-two school-children (twenty-six boys, twenty-six girls), all pupils at local education authority comprehensive schools in London and selected from the average stream of ability as graded by the school. Their age-range was 12 years 10 months to 14 years 6 months. Sixteen subjects were assigned to each of the exclusion instructions and ten subjects to each of the inclusion instructions.

RESULTS

Immediate decoding of 'exclusion' instructions

Fourteen of the thirty-two subjects in the exclusion group, after reading the instruction, responded to the question 'Tell me what you are going to do?' by enumerating the digits which they had to cancel, i.e. these subjects explicitly decoded the instruction in terms of the inclusion class of digits. The remainder of the subjects merely repeated the instruction as given, and may have implicitly decoded the instruction either before commencing the task or during the pre-transfer trials.

The distribution of time scores on the first trial for subjects who were given the exclusion-type instructions was bimodal. A Moses 'test of extreme reactions' (Seigel, 1956) was used to compare the spread of scores of the subjects who explicitly decoded the instruction ($n = 14$) with that of the remainder of the subjects ($n = 18$). A highly significant difference in spread ($P < 0{\cdot}001$) was found and these fourteen subjects were treated as a separate group in the subsequent analyses of performance. This group is referred to henceforth as exclusion group D to distinguish it from the remainder of the subjects who were given the exclusion instructions, exclusion group E. The control group (given the 'inclusion' instructions throughout) is referred to as inclusion group C.

The mean response times on each of the six trials are shown in Fig. 1 for each group. One subject in the inclusion group C misread the instruction for the second set of three trials, and the means for this group are based on nineteen rather than twenty subjects for trials 4–6. Separate analyses of variance were carried out on the average time scores for each group for the pre-transfer trials (1–3) and the post-transfer trials (4–6).

Pre-transfer trials: time scores

The average speed of performance on the pre-transfer trials was 112·6 sec for exclusion group E compared with 72·6 sec for the inclusion group C. The average speed for exclusion group D, the subjects who explicitly decoded the exclusion instruction before the task began, was 80·8 sec. Fig. 1 shows that exclusion group E performed at a considerably slower speed on all three trials than the other two groups. Analysis of variance of the time scores confirmed that the differences between the group means were highly significant ($F = 44{\cdot}9$; D.F. 2, 49; $P < 0{\cdot}005$). Exclusion group E differed considerably from inclusion group C ($P < 0{\cdot}005$) but there was no significant difference between exclusion group D and inclusion group C.

Performance on post-transfer trials

After transfer to the inclusion form of instruction on completion of trial 3, the average speed of performance for trials 4–6 was 88·2 sec for exclusion group E compared with 77·9 sec for exclusion group D and 75·0 sec for the inclusion group C. Fig. 1 shows a marked increase in speed after transfer in the case of exclusion group E, whereas exclusion group D and inclusion group C show little change. Analysis of variance again showed that differences in average speed between the groups were significant ($F = 4{\cdot}2$; D.F. 2, 48; $P < 0{\cdot}01$). Exclusion group E differed significantly from inclusion group C ($P < 0{\cdot}05$). There was no significant difference in average speed of performance between exclusion group D and inclusion group C.

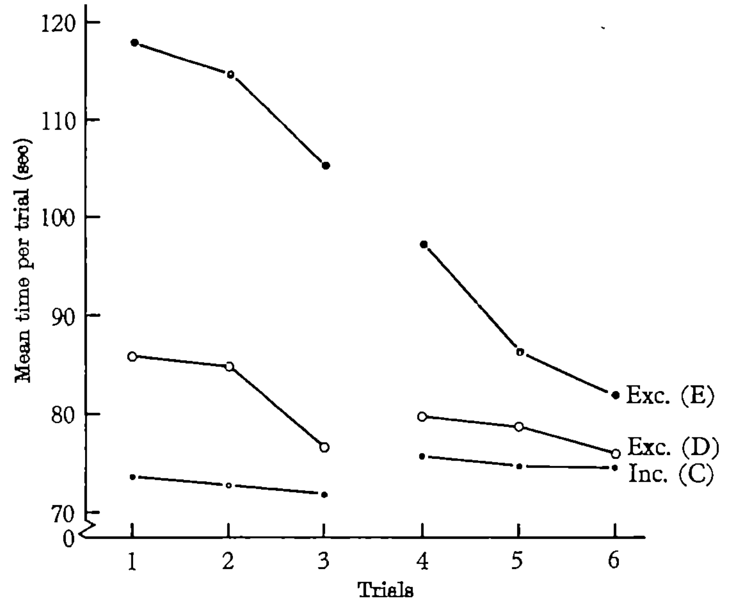


Fig. 1. Transfer as a function of form of instructions in a digit cancelling task. Subjects in exclusion group E, $n = 18$, were given instructions excluding the four digits between '1' and '8', which they were not to cancel, for trials 1–3; subjects in exclusion group D, $n = 14$, were also given these exclusion instructions, but decoded them spontaneously to the equivalent inclusion form before the first of trials 1–3; subjects in the control group, inclusion group C, $n = 20$, were given inclusion instructions which specified the four digits they were to cancel; the subjects in all three groups were given equivalent inclusion instructions for trials 4–6.

Effect of transfer on time scores

The effect of transferring to the inclusion instruction after trial 3 was measured by comparing the difference in mean time scores between trial 3 and trial 4 for the three groups. Exclusion group E made a significant improvement in speed compared with inclusion group C ($P < 0{\cdot}025$). There was no significant difference in change of speed, as a result of transfer, between exclusion group D and the control group.

Errors

The mean percentage of errors made by the three groups for the pre-transfer trials 1–3 (combined) and the post-transfer trials 4–6 (combined) are shown in

Table 1. Two types of error have been distinguished: (*a*) failures to respond to the correct digits (omissions) and (*b*) responses to incorrect digits (false positives).

Exclusion group E made a greater percentage of total errors (14·3 %) than either exclusion group D (5·2 %) or the control group, inclusion group C (6·4 %). Comparing the two types of errors, exclusion group E made a greater proportion of false positives than omissions, whereas for both exclusion group D and inclusion group C the reverse was found. Using the Wilcoxon matched-pairs signed rank tests (Seigel, 1956) a significantly higher proportion of omissions than of false positives was found for both exclusion group D and inclusion group C ($P < 0{\cdot}05$ and $P < 0{\cdot}025$, respectively). The corresponding difference for exclusion group E was not significant.

Table 1. *Mean percentage of errors for the three groups for pre-transfer trials* 1–3 *and for post-transfer trials* 4–6

	Trials 1–3				Trials 4–6			
Group	*n*	False positives	Omissions	Total	*n*	False positives	Omissions	Total
Exclusion (E)	18	8·2	6·0	14·3	18	1·1	4·0	5·2
Exclusion (D)	14	1·4	3·9	5·2	14	1·0	2·6	3·6
Inclusion (C)	20	0·5	5·8	6·4	19	1·1	2·7	3·8

The error patterns for all three groups after transfer were similar, i.e. a greater proportion of omissions than of false positives. The main effect of transfer on error scores was a highly significant drop in the number of false positives made by exclusion group E between trials 3 and 4. (Wilcoxon $T = 0, n = 18, P < 0{\cdot}025$). There was no corresponding change in false positive errors for either of the other two groups. The proportion of omissions made by each group before and after transfer showed no significant change.

DISCUSSION

The evidence on subjects' actual performance when given the opportunity of converting a negatively defined instruction into a positive one must be examined before the findings on subjects' awareness of the potential hazards in handling negative classes can be discussed.

The initial question 'What are you going to do?', which subjects were asked before commencing the task, indicated that fourteen subjects out of thirty-two (44·4 %) had reformulated the instruction explicitly in terms of the inclusion class. The remainder of the subjects repeated the exclusion instruction as given, and there was no evidence at this stage of whether or not they had implicitly decoded the instruction. Their subsequent performance of the task, before and after transfer to the equivalent inclusion instruction, was regarded as crucial evidence. These subjects showed two differences in their performance before transfer from the subjects who used the instruction defined in terms of inclusion. First, they were significantly slower in speed, and secondly, they made significantly more 'false positive' errors. These were the errors made in incorrectly marking the digits specified in the exclusion instruction. Subjects who elected to use the exclusion instruction as given would have to *search* for the digits defined by exclusion but *respond* to the remainder. The tendency to respond (wrongly) to the digits searched for was found in the author's previous experiment (Jones, 1966) for subjects who had no option but to use the exclusion class.

After transfer the same instruction, defined in terms of the inclusion task, was given to all groups, but the response requirements for all subjects remained constant throughout both parts of the task. The only change was the substitution of the equivalent inclusion instruction for the exclusion instruction groups. Those subjects who had explicitly transformed to the equivalent inclusion instruction at the start of the experiment, exclusion group D, showed no difference in speed of performance from the control group, inclusion group C. Those subjects who gave no overt evidence of decoding the exclusion instruction initially, exclusion group E, performed at a significantly slower speed than the control group both before and after transfer to the equivalent inclusion instruction. However, exclusion group E showed a significant drop in false positive errors immediately after transfer, accompanied by a significant increase in speed relative to the control group. After transfer, the pattern of errors for exclusion group E resembled that of both the other groups, i.e. there was a greater proportion of omissions than false positives. The evidence from transfer appears to indicate a change of set in exclusion group E after transfer to the inclusion instructions, i.e. a change in focus from the exclusion class to the inclusion class of digits.

The results indicate that less than half the subjects in the exclusion group (44·4 %) decided to avoid using the negative class of digits by transferring to the inclusion instruction. This suggests that most subjects were unaware of the difficulty of performing a task containing the negative qualifier 'except'. However, the results may not be in conflict with the evidence already cited for preference for using inclusion rather than exclusion classes, and for a dislike of negatively derived inferences. The use of 'except' in a task where only half of the items require a response, as in this experiment, is artificial. In general, it would be assumed that the use of 'except' implied some advantage in terms of memory load. And it would mean that the subjects in the present experiment either did not verify if this was the case, by an initial inspection of the total range of digits, or if they did, decided that the transformation was of no advantage.

Furthermore, an instruction of the type 'Respond to all items except x and y' implies that the task is one of positive response to a large proportion of the items. Only when the subjects commence the task does the conflict engendered by the search for one class of items (the exclusion class) and the response required to the remainder become apparent. The use of the implicit negative 'except' gives a misleading positive set, which leads to a conflation between the search for items in the exclusion class and the positive response to the rest of the items.

Wason's (1965) experiment, on pragmatic factors in language which facilitate response to negative statements, is relevant to this form of instruction. When stimuli are described in terms of an exceptional item and a residual class, statements negating some property of the exceptional item are more readily evaluated than similar statements about items of the residual class. An instruction 'Respond to all items except x and y', implies a similar pragmatic use of the negative, but in the present task it is grossly misleading. In Wason's experiment the exceptionality of the item was indicated by the stimulus array, but not in the linguistic form of the negating statement. It would thus appear that for classification or motor tasks, performed in terms of an exclusion class, the linguistic form of the instruction might well benefit from a more overt expression of the negative. Performance might be expedited if the instruction was expressed in the form 'Do not respond to x and y but respond to the remainder of the items'. This form, although less economical verbally, is less deceptive in indicating the real nature of the task.

REFERENCES

BRUNER, J. S., GOODNOW, J. J. & AUSTIN, G. A. (1956). *A Study of Thinking.* New York: Wiley.

GOODNOW, J. J. & PETTIGREW, T. F. (1956). Some sources of difficulty in solving simple problems. *J. exp. Psychol.* **51**, 385–92.

JONES, S. (1966). The effect of a negative qualifier in an instruction. *J. verb. Learn. verb. Behav.* **5**, 497–501.

SEIGEL, S. (1956). *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill.

WALLACH, M. A. (1959). The influence of classification requirements on gradients of response. *Psychol. Monogr.* **73** (whole no. 478).

WASON, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quart. J. exp. Psychol.* **12**, 129–40.

WASON, P. C. (1965). The contexts of plausible denial. *J. verb. Learn. verb. Behav.* **4**, 7–11.

(*Manuscript received* 22 *March* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 413–418
Printed in Great Britain

PERCEPTUAL DISTORTION INDUCED BY REASONING

BY P. C. WASON AND ADÈLE KOSVINER
Medical Research Council External Staff,
Department of Psychology, University College, London

The aim of this investigation was to compare the effects of three different degrees of expectancy on the perception of relative length. It was assumed that expectancy would be (*a*) maximal when erroneous information about the relative length of two equal lines was *deduced* fallaciously by the subjects, (*b*) intermediate when the same information was *presented* by the experimenter, and (*c*) zero when no such information was presented. It was predicted that the subsequent perception of length would vary in accordance with the degree of expectancy induced, i.e. (*a*) > (*b*) > (*c*). The prediction was confirmed at a high confidence level for the frequency of perceptual judgements, but the amount of perceptual distortion did not differ significantly although the means were in the predicted order. The results are discussed in relation to theories of perception and cognition.

In the last two decades a large amount of research has been done in order to examine the effects of motivational and cognitive factors on perception. The present investigation, however, was designed to test the effects of reasoning on perception. As far as is known, no such experiments have previously been conducted.

The results of previous studies in this area have been discussed by Bruner (1957) in terms of the 'hypothesis', or 'cognitive', theory of perception. It is assumed that in perception discriminatory cues in the sensory input are categorized by an inferential process. The categories used, however, depend on the individual's expectancies about the events to be encountered, and on the search requirements imposed by 'his needs and ongoing enterprises'. A corollary of the theory is that the stronger the expectancy, or 'hypothesis', the greater will be its likelihood of arousal, and its subsequent effects on perception.

There are good grounds for supposing that the processes involved in reasoning arouse a particularly strong degree of expectancy. Reasoning is an internal process which may cost the individual considerable cognitive effort. Furthermore, in deductive reasoning the formal relations between the premises confer logical necessity on the conclusion. Hence the conclusion can be assumed to function as an explicit, directive judgement about the event in question—a judgement which would seem indubitable. In Bruner's terms, the conclusion becomes a 'monopolistic hypothesis'; it admits of no alternative.

The present study, in which expectancy is induced by the subject's own thought processes, is different in certain respects from previous experiments which have investigated the effects of a cognitively supported hypothesis. For example, Bruner & Minturn (1955) have shown that the perception of an ambiguous stimulus varies as a function of the class of stimuli to which the subjects had previously been exposed in a series of training trials. Here the set induced is extrinsic, non-linguistic, and established without cognitive effort. In the present experiment the effects of reasoning were compared with the effects of presenting the conclusion directly. In the former case the subject is actively engaged in making a deductive inference from given premises; in the latter he is the passive recipient of knowledge.

The reasoning process was induced by eliciting a fallacious inference about a conditional statement. Given a statement in the form 'if p then q', two valid and two fallacious inferences can be drawn. It is valid to deduce the truth of q from the truth of p, and to deduce the falsity of p from the falsity of q. It is fallacious, however, to deduce the falsity of q from the falsity of p ('the denial of the antecedent'), and to deduce the truth of p from the truth of q ('the affirmation of the consequent'). The affirmation of the consequent was elicited in the present experiment since previous research has shown that it is committed much more frequently than the denial of the antecedent (Wason, 1964). Only two of the subjects who were tested refrained from making this fallacy, and were accordingly rejected from the sample.

METHOD

Task

Under the two experimental conditions the subjects first of all either inferred, or were told, that one of two equal lines would be longer than the other one. The two lines, which were displayed in the same horizontal plane, were then exposed and the subjects made a perceptual judgement about their relative length. If the line on the left (the variable) was perceived to differ in length from the line on the right (the standard), it was adjusted by the subject until both appeared equal.

Design

There were three independent groups: the reasoning group (RG), in which the subjects inferred that the line on the left would be longer; the information group (IG), in which the subjects were told that the same line would be longer; and a control group (CG), in which no information about the relative length of the lines was given. It was predicted that the frequency of perceptual judgements that the left line was longer, and the amount of its overestimation would vary as a function of the strength of the expectancy induced, i.e. RG > IG > CG.

A sample of forty-five subjects, twenty-one men and twenty-four women, aged between eighteen and twenty-four, was selected at random from the student population of London University, and assigned in rotation to the groups ($n = 15$ in each group).

Apparatus

A strip of red 'sellotape', 0·5 cm wide, was fixed to a board, 63 cm long and 16·5 cm wide. A piece of white cardboard was fixed on top with two slits cut in it to expose two red lines. The line on the right was 15 cm and the line on the left 20 cm in length. The distance separating them was 20 cm. Between the two pieces of board a third piece was inserted under the left-hand slit so that, by moving it in and out, the visible length of the left line could be varied.

Procedure

The subjects were tested individually and were seated at a table in a well-lit room with the apparatus in front of them concealed from their view. The experimenter sat opposite. The line on the left was set at 15 cm, i.e. to be equal in length to the line on the right. The instructions were printed on cards and handed to the subjects at the appropriate phases of the task.

Reasoning-group instructions

'This is an experiment to find out whether perception varies with the amount of information given to people before they look. You will first be given some information about the relative length of two lines, then later on you will be shown these two lines. All the information given to you will be correct. Have you any questions?'

The following conditional statement was then presented for at least 1 min, and for longer if required: 'If the line on the left is longer than the line on the right, then the line on the right is red.'

The statement was then removed and the line on the right was exposed, while the line on the left remained concealed. The following question was then presented to elicit the fallacious inference. 'What can you say *for certain* about the line on the *left*?' (The subjects all replied that the line on the left was longer than the line on the right.)

The line on the left was then exposed and the following instruction presented in order to elicit a perceptual judgement: '*Look at the lines very carefully.* What can you tell me about the length of the line on the left?'

If the lines were reported to be the same length, the experiment was concluded. If the variable was reported to be longer, or shorter, than the standard the following instruction was presented: 'Adjust the line on the left so that it is exactly equal in length to the line on the right.' The amount of adjustment was measured to the nearest 0·1 cm.

Finally, the following question was presented to elicit an introspective report. 'Tell me, did the line on the left actually *look* longer to you, or did you think it *must* be longer?'

Information-group instructions

The general instructions were the same as those given to the reasoning group, but the subjects were not told that all the information given to them would be correct. The following instruction was then presented. 'The line on the left is longer than the line on the right.' Both lines were then exposed and the subjects instructed to make a perceptual judgement, followed if necessary by an adjustment of the variable and an introspective report, as in RG.

Control-group instructions

The general instructions were the same as those given to the reasoning group, but the subjects were told that they would receive no information. Both lines were then exposed, and the subjects were instructed to make a perceptual judgement, followed if necessary by an adjustment of the variable, as in RG. The request for an introspective report was omitted.

The responses were not timed in any of the groups as it was considered that this might have induced stress or impatience; in a pilot study the mean time taken to make a perceptual judgement was 40 sec.

RESULTS

Table 1 shows the number of subjects in each group who perceived the variable as longer than or equal in length to the standard. Only three subjects, one in each group, perceived it as shorter, and they were included in the equal category.

Table 1. *Frequency of perceptual judgements about the relative length of the variable line* ($n = 45$)

	RG	IG	CG	
Longer	12	4	1	17
Equal	3	11	14	28
	15	15	15	45

The prediction that perceptual judgements of relative length would vary as a function of the degree of expectancy induced, i.e. $RG > IG > CG$, was highly significant ($S = 330$, $P = 0{\cdot}00005$, one-tailed). However, it will be noted that this trend is almost entirely due to the difference in the relative frequencies between RG and IG, the difference between the frequencies in IG and CG being negligible. There is evidently a large difference between inferring a conclusion and being told it, but the latter hardly affects perceptual judgement at all.

The mean amount of overestimation of the length of the variable was + 0·82 cm in RG ($n = 13$), + 0·62 cm in IG ($n = 5$), and zero cm in CG ($n = 2$). Although the means were in the expected order, the corresponding prediction for the amount of overestimation was not confirmed.

Eleven out of twelve subjects in RG, and three out of four in IG, who initially perceived the variable as longer, subsequently reported with different degrees of emphasis that it actually looked longer. However, no correlation was apparent between the emphasis of the introspective report and the amount of adjustment made. For example, the subject in RG who made the largest adjustment (1·8 cm) reported subsequently: 'It looked slightly longer, couldn't be sure'; and a subject in the same group who made a relatively small adjustment (0·3 cm) reported emphatically: 'The left line *is* longer'. Similarly, the subject in IG who made the largest adjustment (1·7) cm reported subsequently that the lines looked the same length.

It will be noted that there was a slight discrepancy between the initial instructions given to RG and IG. In the former group, but not in the latter, the subjects were told that all the information given to them would be correct. It was observed in the pilot study that unless this point was stressed the subjects seemed to give scant attention to the conditional statement. Hence in the experiment the sentence, 'all the information given to you will be correct', was included in the instructions to RG. It was omitted in the instructions to IG, however, because it seemed redundant, as well as being false. This small difference could have confounded the results. Reinforcing the truth of the conditional statement might have *artificially* accentuated the presumed truth of inference derived from it. Hence it was decided to check on this factor by testing another IG in which the relevant instruction about the correctness of the information was equated to that given in RG.

A further sample of fifteen subjects, nine men and six women, was randomly selected from the same population. The results showed that the relative frequency of perceptual judgements was identical to that obtained in the original IG, the mean overestimation of the length of the variable being + 0·78 cm ($n = 5$). The subsequent introspective reports, however, revealed a conflict between the information presented and the perception of relative length, e.g. 'I assumed it was longer, but it looked the same length to me'; 'They looked the same—then it looked longer after I thought it must be, so I looked again to see'; 'I'm very credulous—I'm an American—it looked a little longer to me, or it may be the same, but I say it's longer.'

This subsidiary result suggests that the reinforcement of false information as correct tended, if anything, to create greater doubt rather than greater certainty. It can be concluded that the difference in the original instructions to RG and IG was unimportant.

DISCUSSION

The prediction that the perceptual judgements about relative length would vary as a function of the degree of expectancy induced was confirmed at a high confidence level. This result corroborates the 'hypothesis' theory in a new way. It strongly suggests that the effort and ego-involvement which occur in reasoning, and the explicit inference which terminates it, produce a marked distortion of the stimuli in accordance with the subjects' hypothesis. (The fallaciousness of the reasoning is presumably an irrelevant factor.) Conversely, the passive acceptance of the information contained in the conclusion arouses a weaker hypothesis which is readily infirmed by reality. Indeed, the surprising aspect of the results is that the frequency of perceptual judgements do not appear to differ significantly between IG and CG. It is evident that the present attempt to induce expectancy by direct suggestion had a negligible effect on perception.

The fact that the results were obtained with the present task appears to strengthen the theory. Previous experiments have all used 'marginal' conditions of stimulus presentation in order to facilitate non-veridical perception, e.g. tachistoscopic exposure, 'blurred pictures', ambiguous figures, and so on. In this experiment the subjects were allowed unlimited time to inspect the material before making a perceptual judgement, and the only concession to non-veridical perception was that the distance between the two lines (20 cm) precluded the comparison of their relative length with a single fixation of the eyes.

The prediction was not confirmed for the amount of overestimation of the variable. This result could be taken at its face value, i.e. the amount of perceptual distortion does not differ as a function of the degree of expectancy induced. It seems quite likely, however, that the subjects' adjustment of the variable was not sufficiently accurate to reflect their perception of relative length objectively. More sensitive techniques of adjustment will be used to investigate this factor in subsequent experiments.

The results of this experiment are compatible with a quite different theoretical interpretation in terms of cognitive dissonance (Festinger, 1957; Brehm & Cohen, 1962). The expectancy induced may have had no perceptual effects and the introspective data may have been rationalizations. The subjects may well have been uncertain, in the first instance, whether the two lines were different in length. But in RG they had committed themselves to an overt inference that the left line was longer; and they had done so publicly in a social situation. Under these conditions a judgement that the lines, in fact, were equal in length would have forced the individual to contradict himself, and this might have entailed a loss of self-esteem. Hence the dissonance would have been resolved by interpreting a relatively ambiguous reality in accordance with the inference. (It is an open question whether a covert inference would have had similar consequences.) This interpretation of reality need not have involved deliberate self-deception. The subjects might have persuaded themselves, on looking at the lines very hard, that the left one was longer. In IG, on the other hand, the subjects had not committed themselves to a cognition—they had merely entertained one. Hence the information presented by the experimenter might have been rejected more readily in the face of an incongruous reality represented by two apparently equal lines.

It is difficult to evaluate the force of this explanation. Since percepts are phenomenal data it is impossible to assess the extent to which the present results are influenced by perceptual or cognitive processes. Bruner (1957) has argued cogently that the two inter-act in ordinary 'perceiving' in everyday life. But whatever sort of explanation is more plausible, the results lend some small empirical support to the thesis that a belief, to which an individual is committed, may distort his perception of reality, or at least remain immune to correction by experience. In the present task, however, the belief is not a value judgement or a prejudice in which emotion might have been invested in the interests of adaptation. It is fallacious inference about a trivial event. But perhaps the cognitive processes which culminate in judgement are more important for the preservation of a belief than its content. Or does the act of judgement itself make a belief resistant to extinction?

REFERENCES

BREHM, J. W. & COHEN, A. R. (1962). *Explorations in Cognitive Dissonance*. New York: Wiley.

BRUNER, J. S. & MINTURN, A. L. (1955). Perceptual identification and perceptual organization. *J. gen. Psychol.* **53**, 21–28.

BRUNER, J. S. (1957). On perceptual readiness. *Psychol. Rev.* **64**, 123–52.

FESTINGER, L. (1957). *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.

WASON, P. C. (1964). The effect of self-contradiction on fallacious reasoning. *Quart. J. exp. Psychol.* **16**, 30–4.

(*Manuscript received* 14 *June* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 419–421
Printed in Great Britain

A REPLY TO McPHERSON'S COMMENTS ON THE ABERDEEN CONDITIONING STUDIES

BY H. J. EYSENCK
Institute of Psychiatry, University of London

Mr McPherson puts us in his debt by drawing attention to the omission on the writer's part (Eysenck, 1965) of mention of his work on the relation between eyeblink conditioning and personality (McPherson, 1965). Interesting replications of major experiments, such as his, ought to be published, if only in brief; otherwise it is difficult to obtain knowledge of their existence. When the writer inquired after relevant studies on conditioning and personality, he wrote to Prof. Fraser, of Aberdeen, on 16 April 1964, stating his intention of writing up the literature and asking to borrow a thesis on this subject which he believed to have been completed there; Prof. Fraser kindly sent along the theses of Brebner (1957) and Symon (1958). No mention was made of McPherson's work, of which therefore the writer remained in ignorance. This is no doubt only one of many such omissions; another one which has been traced recently is an M.A. thesis by Jacobs (1965). In this he compared fifteen hysterics with a mean M.P.I. extraversion score of 25·4 with fifteen dysthmics, mean E score 13·7, on eyeblink conditioning. The hysterics gave a mean number of 7·33 conditioned responses during acquisition, and 2·2 during extinction; dysthymics gave 13·06 and 4·93 respectively. Both differences were significant beyond the 1% level. These results are more in line with those of Brebner and Symon (and of Franks, 1956, 1957) than with those of McPherson, and although Jacobs's method of recording appears to have been less satisfactory than that of the Aberdeen group his results may be of some interest in showing that McPherson's results are 'the odd man out' and require explanation for his failure to duplicate successfully well-established findings.

McPherson also draws attention to various features of the Brebner and Symon studies which, in the brief summary made by the writer, had not been mentioned. Thus it is true that Brebner found a sex × personality interaction; Symon, on the other hand, did not. It is not clear how far this is relevant to the theory under discussion; in any case, the significant interaction with sex in Brebner's study was in fact mentioned (Eysenck, 1965, p. 260), as McPherson acknowledges. He is right in saying that such an interaction is not predictable from the writer's theory, but, as it was not replicated in Symon's parallel study, and was based on very few cases, it would not really seem to call for an explanation until shown more clearly to be more than an accidental and non-repeatable artifact. McPherson also points out that in the Brebner study three subjects, all introverted, had been excluded because they failed to condition. The writer has not put forward any theory to account for the occasional non-conditioning subject; it seems likely that these are in some qualitative sense different from those who do condition, if only with difficulty, and consequently it is doubtful if mention of this particular fact would have been very relevant.

McPherson transformed the raw scores in Symon's study to logarithms 'since the requirement of homogeneity of variance was not satisfied'. He found on these transformed scores that the significance of the differences between extraverts and introverts disappeared. This very common practice is not sensible from a mathamatical point of view (Scheffé, 1959); as Box (1953) has pointed out, the F test is not sensitive to failure of variance inequality, whereas Bartlett's test for variance homogeneity is! The whole issue is discussed with great clarity and force by Cohen (1965) who makes clear the reasons why McPherson's procedure is inadmissible. The writer would go as far as to say that, if a transformation had to be made, then antilogs should have been used; the reasons for this recommendation are psychological, not mathematical, and follow from the argument in Eysenck (1965), footnote 5. Altogether, it would seem injudicious to use transformations for pseudo-statistical reasons in disregard of contrary psychological reasons, and little serious attention should be paid to the conclusions drawn from such computations.

McPherson throughout his brief note seems to suggest that the onus for explaining the departures from his theoretical predictions which occur in the Aberdeen work lies on the present writer. This seems an odd view. Two of the three Aberdonian studies agree at a reasonable level of significance that introverts condition better and quicker than extraverts; in this they support the earlier work of Franks on normal and neurotic subjects. Combining, as we may, the probabilities of these studies gives us an overall probability against the null hypothesis of something better than $P < 0{\cdot}000001$. McPherson's failure to obtain results in line with those of his fellow Aberdonians, or with Franks, should lead him to search for sources of error in his procedure, or in any case for differences which might explain the discrepancy between his results and theirs—not between his results and the writer's theory. There are many unreported aspects of such experiments which could lead to differences in outcome, and no one from the outside could be expected to suggest why McPherson failed to obtain results, under apparently identical circumstances, similar to those of Brebner and Symon. It is curious that McPherson should regard this as a failure, not of himself, but of the writer's theory; a somewhat less extrapunitive attitude might have led him to discover the reasons for the discrepancy, and thus make a genuine contribution to the control of extraneous factors in eyeblink conditioning. His results and arguments certainly do not justify his claim that 'the Aberdeen work appears to show little evidence that introverts are more conditionable'.

It is also difficult to see the force of his suggestion that 'an explanation of [the] Aberdeen findings in terms of anxiety aroused by the experimental situation was suggested by Reid (1964)'; no mention of such an explanation (or even of anxiety) was found in the source quoted. In any case, neuroticism/anxiety scores were available to the Aberdeen investigators, and it would have been very simple for McPherson to run a correlation between conditioning and N; past work has nearly always shown such correlations to be insignificant unless special steps were taken to make the conditioning situation specially anxiety arousing (Spence, 1964). Inspection of such data from Aberdeen as were available to the writer did not suggest that they would prove an exception to this rule.

REFERENCES

Box, G. E. P. (1953). Non-normality and tests of variance. *Biometrika* **40**, 318–35.

Brebner, J. M. T. (1957). An experimental investigation of the relationship between conditioning and introversion–extraversion in normal subjects. Unpublished M.A. thesis, University of Aberdeen.

Cohen, J. (1965). Some statistical issues in psychology research. In *Handbook of Clinical Psychology*, ed. B. B. Wolman. London: McGraw-Hill.

Eysenck, H. J. (1965). Extraversion and the acquisition of eyeblink and GSR conditioned responses. *Psychol. Bull.* **63**, 258–70.

Franks, C. M. (1956). Conditioning and personality: a study of normal and neurotic subjects. *J. abnorm. soc. Psychol.* **52**, 143–50.

Franks, C. M. (1957). Personality factors and the rate of conditioning. *Brit. J. Psychol.* **48**, 119–26.

Jacobs, M. (1965). Conditionability: a clinical investigation of the bi-dimensional theory and classification of the neuroses. Unpublished M.A. thesis, University of the Witwatersrand.

McPherson, F. M. (1965). Comment on Eysenck's account of some Aberdeen studies of introversion–extraversion and eyeblink conditioning. *Brit. J. Psychol.* **56**, 483–4.

Reid, R. L. (1964). Conditionability and eyeblink conditioning. *Bull. Brit. psychol. Soc.* **17**, no. 55, 16A.

Scheffé, H. (1959). *The Analysis of Variance*. New York: Wiley.

Spence, K. W. (1964). Anxiety (drive) level and performance in eyelid conditioning. *Psychol. Bull.* **61**, 129–39.

Symon, S. M. (1958). An investigation of the relationship between conditioning and introversion–extraversion in normal subjects. Unpublished M.A. thesis, University of Aberdeen.

(*Manuscript received* 17 *January* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 423–429
Printed in Great Britain

COMMENTARY ON 'THE GENETIC DETERMINATION OF DIFFERENCES IN INTELLIGENCE: A STUDY OF MONOZYGOTIC TWINS REARED TOGETHER AND APART' BY CYRIL BURT: CONGENITAL INFLUENCES ON THE DEVELOPMENT OF TWINS

BY D. H. STOTT
*Department of Psychology, University of Glasgow**

Burt (1966) has contributed the most comprehensive study of twins reared apart which has been made so far. His results, founded upon a methodology which has enabled him to overcome many of the former objections, leave no doubt that there is some genetic reason why the intelligence test scores of twins tend to resemble each other much more closely than do those of siblings. This conclusion is inescapable for a parsimonious interpretation of his results.

However, Burt's own conclusion, that 'individual differences in "intelligence"... are influenced far more by genetic constitution, or what is popularly termed "heredity", than by post-natal or environmental conditions' (p. 152), involves a number of other assumptions, and the acceptance of a particular conceptual framework, which may be questioned. He proposes that there is some distinctive cognitive ability or quality termed 'intelligence' which largely accounts for individual differences in test scores. He juxtaposes hereditary and post-natal influences as if the prenatal phase of human development were unimportant. By using the term 'more' he implies that these postulated influences are sufficiently independent to permit of quantitative assessment. And he assumes that the 'contributions' of these genetic and (postnatal) environmental factors are fixed irrespective of the population studied; that is to say, that it is possible to extrapolate quantitatively from studies of twins to populations mainly composed of singletons. Each of these issues demands monograph treatment. The more limited objective of the present commentary is to argue that how Burt's results are interpreted depends upon the particular conceptual framework into which they are fitted.

Burt repeats his profession of faith in an 'innate, general, cognitive factor' which is separate from conation or motivation. It is difficult to reconcile this concept with the developmental view of intelligence that is now generally held. It is a fact of observation that individual children show varying success in solving certain standardized problems. But it is not necessary to attribute this variation to differences in some cognitive factor. No problem can be solved by pure 'intelligence'. A person has to know about the nature and qualities of all the things that he is dealing with, and such knowledge comes in large part from experience. Even apart from the advantages enjoyed by children in different cultural groups, the extent to which an individual learns about his environment depends upon his curiosity, his eagerness to understand, and his habits of cogitating upon his observations. In other words, his contemporary problem-solving ability is a function of the strength of his intrinsic motivations during the earlier stages of his development. Developing the argument further, the ability to solve problems of a given type depends upon the acquisition of a stock of trial solutions or 'mental templates' which are applied as analogies from similar, previously solved problems. Their possession by any individual depends upon the extent to which he has, up to that point, engaged himself in problem-solving. This in turn depends upon the strength of his 'effectiveness-motivation' (Stott, 1961), and his choice of problem-solving as a sphere of effectiveness. A child without the confidence or desire to widen the scope of his effectiveness will remain at a low level of *cognitive* development.

---

* Now at Department of Psychology, University of Guelph, Ontario, Canada.

In short, the conceptual equipment of any individual, upon which his problem-solving ability and success at intelligence tests depend, is in large part determined by his motivational history. This background is not open to observation at any single point in the course of mental development; hence it is easy to slip into the assumption that A simply has more of an innate general cognitive factor than B. But it could equally well be argued that the main determinants of differences in intelligence-test scores are motivational, and observation of the motivational deficiency and disorganization of most mental defectives supports this view. There is nothing in Burt's results, for example, to contradict the hypothesis that the operative genetic factor making identical twins more similar in 'intelligence' is a motivational one.

At all events, the means are lacking to distinguish the influence of motivation from differences in a hypothetical cognitive ability. There is no convenient phenotype to present to the geneticist such as the colour of eyes or hair. All the data are behavioural. Hence, Burt's first tenet, that the genetic factor is 'intelligence' or a general cognitive ability distinct from conation, does not emerge as a necessary conclusion from his findings.

Burt's second major assumption—that the 'contributions' of heredity and (post-natal) environment are relatively constant irrespective of the population studied—is implicit in his general conclusion. Nevertheless, it is plain that the range of environmental contrast will affect the degree to which 'environment' influences mental development, and that there are wide variations in the homogeneity of different environments. Gordon (1923) found the so-called environmental factor to be decisive in his study of canal-boat children who were presented with test tasks culturally quite foreign to them. The children were near-feebleminded judging by the scores, but their parents earned more than the skilled workers or young teachers of their time. In the Mooseheart study (Raymer & Hinton, 1940), on the other hand, the environmental factor in the form of educational enrichment proved nugatory.

Burt's third major premise, that inferences about the extent of a genetic influence in twin development can be extended to non-twins, has been challenged by Zazzo (1960), to which important work however he makes no reference. Zazzo's thesis is that twins are so atypical developmentally that no quantitative conclusions about general genetic–environmental influences can be drawn, and some of his evidence is summarized below. The critical issue is the sort of atypicality and, more specifically, whether the genetic determinants act to the same degree in twins and singletons.

If they are of relatively fixed expressivity and are independent of the influence of intra-uterine environment, Burt's extrapolation of his results on to the human race in general would have some plausibility. But the weight of evidence in experimental biology is against him. A large number of studies, including the classic studies of Landauer & Bliss (1946) and of Fraser, Kalter, Walker & Fainstat (1954), demonstrate that the genetic factors making for defect are in the nature of *potentialities* for variation rather than genes of constant expressivity. The potentiality is triggered off or facilitated, to use the term of Malpas (1937), by gestational stress. Several workers (e.g. Thompson, 1957; Ader & Belfer, 1962) have shown that the same applies to variations in 'emotionality' or 'motivation'. Werboff, Broeder, Havlena & Sikov (1961) found that behavioural dysfunction was related to stress in the later stages of pregnancy, a result that is paralleled by the writer's finding that 'unforthcomingness', or 'effectiveness-deficit' as the condition might more accurately be termed, is also closely related to stress late in pregnancy (Stott, 1959*a*).

Burt discounts genetic–environmental interaction at the uterine stage as a reason for the greater similarities of monozygotic twins. His argument is directed to showing that, in spite of sharing the same pregnancy, and in the case of monozygotic twins the same membrane, there would invariably be differences in the conditions of prenatal development for each twin. This point is not in question, but it is not the main point. If twinning in itself reflects a liability to gestational stress, it is to be expected that the genetic potentialities for impairment will be triggered off more frequently among twins than among singletons.

The evidence is that twins are vastly more affected by prenatal stresses than are singletons. In Zazzo's study of 718 twins their period of gestation was on average 20 days shorter and their mean birth weight some 800 g less; moreover, mean birth weight was significantly lower for monozygotic than for dizygotic twins ($P < 0{\cdot}001$). Compared with singletons, perinatal mortality was twice as high among dizygotic twins and no less than eight times as high among monozygotic twins. These different degrees of impairment for the two kinds of twin call into question the whole method of twin study in so far as it is used to extract quantitative estimates of the relative effects of heredity and environment. Apart from the perinatal period, Zazzo points out that mortality among twins remains higher at least up to the age of three years. The writer's study of children who had suffered early illness showed to what extent the presumably congenital impairment takes on a multiple character and affects mental development (Stott, 1959*b*; summarized in Stott, 1966).

There is evidence that the tendency to congenital impairment has a general effect on the mental development of twins. In Zazzo's study they showed a deficit of 5–8 points in test intelligence, and in the Scottish Mental Survey (Scottish C.R.E., 1953) a deficit of nearly 5 points. In Wiseman's recently published Lancashire survey the mean intelligence test score of the twins was only 20·81 compared with 32·20 for the total sample (Wiseman, 1964). This massive deficit of 11·39 points may reflect a higher survival rate of multiply impaired twins owing to the introduction of antibiotics and other improved methods of medical treatment. If the latter explanation holds, it may be predicted that future samples of twins will continue to show a similarly great intellectual inferiority, and will not be comparable even with earlier twin samples, let alone singletons.

The implication of all these findings is that the mental inferiority of twins must be attributed to the gestational risks of twinning, since there can be no genetic reason why twins should be so affected. Further evidence of this congenital, as distinct from purely genetic factor, is seen in the greater relative inferiority of boy twins: in the Scottish survey they averaged 3·06 points less than boy singletons, compared with a relative inferiority of only 1·77 points for the girls. Like-sexed twins generally, including monozytogic twins, have tended to be mentally inferior to twins of different sex. In Zazzo's study speech defect and retardation showed the same pattern: these conditions were considerably more frequent among boy twins, and also among monozygotic compared with dizygotic twins, whether boys or girls. The percentages of combined speech defect and retardation were:

	Monozygotic twins	Dizygotic twins
Girls	11·3	7·3
Boys	16·6	6·3

This picture of a greater impairment in males is so characteristic of congenital conditions (excluding certain malformations of the c.n.s.) that it can be used as a criterion of congenitality (Stott, 1966). It is noteworthy that Burt makes no reference either to the general mental inferiority of twins or to the relative inferiority of the males, and in the present study he does not give the mean I.Q. of his twin groups, either as a whole or by sexes.

The present thesis is, in sum, that there is a third explanation of individual human differences other than in terms of the usual genetic–environmental dichotomy. It is that within all species, owing to the need to control population numbers (Stott, 1962), there is genetic provision for multiple impairments which normally remain unexpressed as mere potentialities, but which are triggered off by those gestational stresses which would be associated with high population density (malnutrition, the fatigues of long treks, distress and harassment arising from invasion of territory, heightened exposure to virus infections, and so on. The resulting impairments tend, in extreme cases, to decimate the juvenile population, and in less extreme cases to reduce the efficiency of a significant proportion (in man by mental subnormality, behaviour disturbance, motor impairment, physical defect, reproductive failure). The cutting down of a population would be less detrimental to future recovery if males were culled out more drastically than females, hence the greater liability of males to congenital impairment.

This 'congenital' explanation postulates an interaction of a special sort between genetic and environmental influences. It could account for the greater mental similarity of monozygotic twins in that each of the pair would be endowed with the same genetic potentiality for defect (whether this be of a motivational or of a cognitive structure); it could possibly also account for the occasional supranormal motivation which seems to arise from the same conditions and which may lead to abnormally high or unusual mental development (hence genius and its attendant ills). Moreover, by sharing the same pregnancy, twins also share the same hazards to maternal health and emotional state. This would explain the somewhat higher correlation between fraternal twin pairs compared with siblings, to which Vernon (1960) has drawn attention as an unsolved problem.

It remains to be asked whether this congenital chain of events could account for a significant part of the observed variability. Again there is evidence to which Burt does not refer. In his study of monozygotic twins of school age Shields (1954) found that the sexually more developed were more intelligent in the ratio of 21:10. Among the adult female monozygotic pairs of his later study (Shields, 1962), those who had begun to menstruate at an earlier age were more intelligent in the ratio of 27:7. He points out that with $\chi^2 = 10{\cdot}618$ ($0{\cdot}005 > P > 0{\cdot}001$) this was his most significant finding as regards test intelligence. The first-born were also more intelligent in the ratio of 26:17. The differences in intelligence reflected in these ratios suggest a general tendency to a multiple congenital impairment. It would seem that the congenital (genetic + intra-uterine + delivery) factor is a massive rather than a marginal one.

From the full case-studies which Shields (1962) has given for his forty-four adult twin pairs, the writer has been able to analyse in detail to what extent the major discrepancies in intra-pair test intelligence could be accounted for on the hypothesis of multiple congenital impairment. Shields's own criteria of discrepancy were used: 15 or more points on the combined score, or corresponding differences in the separate tests. Of the eighteen pairs in which intelligence was discordant, Shields himself adduces physical causes which were probably congenital in nine, to which must be added one case of congenital syphilis. Of the remaining cases important congenital factors can be discerned in five, while no data are given for the remaining three. Thus a congenital reason for the discrepancy is probable in fifteen of the eighteen discordant cases. Newman, Freeman & Holzinger (1937) did not give systematic data on the early lives of their nineteen separated monozygotic pairs, only mentioning ill health or the reverse when it was marked. Nevertheless, an examination of their case-histories shows that in the three cases with an I.Q. discrepancy of 12 points or more it is the unhealthy or delicate twin who had the lower score.

There is another implication of the 'congenital' hypothesis for the study of twins. The reasons for separating pairs of twins imply a greatly heightened probability of gestational stress, and constitute a powerful selective factor. Among Shields's forty-four pairs, twenty were separated because the mother was unable to manage both twins (owing to physical or mental illness, or to poverty), ten because the mother died at the birth, seven because of illegitimacy, five because the father was dead or sold one of the twins to settle his debts, while two were taken into public care on account of home conditions. These reasons are paralleled in Burt's sample. In short, if powerful determinants operate at the prenatal stage, separated twins are more likely to be affected than twins not separated. Hence separated monozygotic twins are not congenitally typical of monozygotic twins in general.

With these complications any resemblance to a Mendelian pattern of inheritance, of presumed constant expressivity, that Burt discerns carries little conviction. Many earlier attempts have been made to establish regular genetic patterns from the results of twin studies. In an authoritative review of work up to the time Price (1950) made the suggestion that 'the genes and modifiers were such that the anomaly concerned would not have been expressed ordinarily, but imbalance in the mutual circulation occasioned expression of the character in only one twin'.

It is noteworthy that, on the evidence of twin studies, Price arrived at an independent formulation of the theory of facilitation, albeit in relation to a particular sort of gestational stress. Moreover, he quoted several instances of malformations and other defects occurring in only one of a monozygotic pair, including one case of discordance for colour blindness. Professor Pickford and the writer have been able to find four further authenticated cases in the literature (Pickford, 1966); since they were found fortuitously in general samples, among whom twins would number perhaps only one in 100, such discordance may prove to be comparatively common. In all five pairs there was a family history of colour blindness. Presumably, the expressivity of the recessive gene (in heterozygous form in the affected females) would have been modified by differences, as between one monozygotic twin and the other, in the inter-uterine environment. Major defects of colour vision have been regarded as conforming particularly well to a regular Mendelian expressivity; if the true state of affairs is one of a genetic potentiality activated by gestational stress in their case, what hope can there be of finding that the varied assortment of behaviour which gives rise to the concept of 'intelligence' should be inherited in any uncomplicated and demonstrable way?

There are other reasons why it is difficult to accept the tenuous resemblance to a Mendelian pattern which Burt adduces. If parent–child correlations in 'intelligence' are in the main due to inheritance, supporting evidence should emerge from studies of children in foster homes. The most notable finding from these studies and from follow-up studies of the children of mentally defective mothers (Brandon, 1957), is that low 'intelligence' at any rate is not inherited. The small parent–child correlations which have been systematically reported (Burt gives no particulars of his finding of one of 0·5) can well be attributed to the tendency, noted by several investigators, for adopting agencies to place infants in foster homes of comparable socioeconomic level to those of the true parents, presumably out of a sense of genetic justice. If 'intelligence' is inherited in the massive and regular way which Burt suggests, he has missed an opportunity of demonstrating the point for the sixteen twins of professional and clerical parents in his sample who were fostered into manual workers' families (cf. Burt, 1966, Table 1).

References

Ader, R. & Belfer, M. L. (1962). Prenatal maternal anxiety and offspring emotionality in the rat. *Psychol. Rep.* **10**, 711–18.

Burt, C. (1966). The genetic determination of differences in intelligence: a study of monozygotic twins reared together and apart. *Brit. J. Psychol.* **57**, 137–53.

Brandon, M. W. G. (1957). The intellectual and social status of children of mental defectives. *J. ment. Sci.* **103**, 710–38.

Fraser, F. C., Kalter, H., Walker, B. E. & Fainstat, T. D. (1954). Experimental production of cleft palate with cortisone and other hormones. *J. cell. comp. Physiol.* **43** (suppl. 1), 237–59.

Gordon, H. (1923). *Mental and Scholastic Tests among Retarded Children.* Educ. Pamphlet no. 44. London: H.M.S.O.

Landauer, W. & Bliss, C. I. (1946). Insulin-produced rumplessness of chickens. *J. exp. Zool.* **102**, 1–22.

Malpas, P. (1937). The incidence of human malformations and the significance of changes in the maternal environment in their causation. *J. Obstet. Gynaec. Brit. Emp.* **44**, 434–54.

Newman, H. N., Freeman, F. N. & Holzinger, K. J. (1937). *Twins: a Study of Heredity and Environment.* Chicago University Press.

PICKFORD, R. W. (1966). Variability and consistency in the manifestation of red–green colour vision defects. *Vision Res.* (in the Press).

PRICE, B. (1950). Primary biases in twin studies. *Amer. J. hum. Genet.* **2**, 293–352.

REYMERT, M. L. & HINTON, R. T. (1940). The effect of a change to a relatively superior environment upon the I.Q.'s of 100 children. In *39th Yearbook, Nat. Soc. Study Educ.*, Bloomington, Ill.

SHIELDS, J. (1945). Personality differences and neurotic traits in normal twin schoolchildren. *Eugen. Rev.* **45**, 213–46.

SHIELDS, J. (1962). *Monozygotic Twins Brought up Apart and Brought up Together*. London: Oxford University Press.

STOTT, D. H. (1959*a*). Evidence for prenatal impairment of temperament in mentally retarded children. *Vita Humana* **2**, 125–48.

STOTT, D. H. (1959*b*). Infantile illness and subsequent mental and emotional development. *J. genet. Psychol.* **94**, 233–51.

STOTT, D. H. (1961). An empirical approach to motivation based on the behaviour of a young child. *J. child Psychol. Psychiat.* **2**, 97–117.

STOTT, D. H. (1962). Cultural and natural checks to population growth. In *Culture and the Evolution of Man*, ed. M. F. A. Montagu. New York: Oxford University Press.

STOTT, D. H. (1966). *Studies of Troublesome Children*. London: Tavistock.

THOMPSON, W. R. (1957). Influence of prenatal maternal anxiety on emotionality in young rats. *Science*, **125**, 698–9.

VERNON, P. E. (1960). *Intelligence and Attainment Tests*. London: University of London Press.

WERBOFF, J., BROEDER, J. D., HAVLENA, J. & SIKOV, M. R. (1961). Effects of pre-natal x-ray irradiation on audiogenic seizures in the rat. *Exp. Neurol.* **4**, 189–96.

WISEMAN, S. (1964). *Education and Environment*. Manchester University Press.

ZAZZO, J. (1960). *Les Jumeaux : le Couple et la Personne*. Paris : Presses Universitaires de France,

(*Manuscript received* 27 *July* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 431–433
Printed in Great Britain

COMMENTARY ON 'THE GENETIC DETERMINATION OF DIFFERENCES IN INTELLIGENCE: A STUDY OF MONOZYGOTIC TWINS REARED TOGETHER AND APART' BY CYRIL BURT

BY D. G. LEWIS
Department of Education, University of Manchester

In the paper to which Burt (1966) refers (Lewis, 1957), mention was made of a multifactorial model for the genetic determination of intelligence, a model which is oversimplified (no account being taken of partial dominance, assortative mating and linkage) but which leads to a normal distribution. The main purpose of the paper, however, was to point out that the normal distribution of intelligence, defined operationally by tests such as the Stanford–Binet, is not a 'brute fact' of nature, but rather a conceptual matter, though one supported by plausibility, mathematical convenience and common-sense. Thus the approximately normal distribution of Binet I.Q. appears as a direct consequence of the statistical criteria used to select the test items at each of the age levels; the fact that Binet and his successors did not *deliberately* select for such a distribution is irrelevant. It is reasonable to add that the normal distribution of 'genotypic intelligence', the 'Intelligence A' of Hebb (1949), is no more than an assumption, necessarily untestable and of little practical importance, except in so far as there may also be reason to assume that environmental conditions do not modify this distribution, or modify it only in certain specified ways.

The core of Burt's present paper is the set of correlations obtained from monozygotic twins separated at an early age, with the claim that these correlations derive from the common heredity of the twins rather than from similarity in their subsequent environments. That the environments were fundamentally different needs to be assured. The objection that environments were similar (through selective placements) has been anticipated by a breakdown of the occupational status of both parents and foster parents (cf. Burt, 1966, Table 1). Burt maintains that these figures should dispose of the explanation that the high correlations for the separated twins are due to the way the foster parents were chosen. In this however he is clearly too optimistic.

It is true that few of the twins were placed with foster parents of precisely the same social class as their parents. Many, however, were placed with foster parents of a very similar social class, i.e. in adjacent categories. This is evident if the groupings are broadened by combining categories I and II (as one 'professional' class), categories III and IV (as a 'clerical and skilled manual' class), and categories V and VI (as a 'semi-skilled and unskilled manual' class). These combinations, omitting the numbers in the residential institution, give the figures in Table 1. In over 40 % of the cases, nineteen out of forty-six) the twins were in fact separated in socially similar homes (compared with only nine twins separated in homes of exactly the same class). Evidently the endeavours of social workers to place twins in homes of a broadly similar class to that of their parents had met with at least a fair measure of success.

Burt would undoubtedly strengthen his case if he could produce separate sets of intra-class correlations for monozygotic twins separated in socially similar homes and for monozygotic twins separated in socially dissimilar homes, and if the correlations in the former set were not markedly higher than those in the latter. It is unlikely that any statistically significant difference would emerge from the relatively small samples at present available, but the *direction* of the differences, unless the differences for both group and individual tests proved negligibly small, might still be revealing.

Table 1. *Occupational categories of parents and foster parents, I and II—'professional,' III and IV—'clerical and skilled manual', V and VI—'semi-skilled and unskilled manual'*

	Foster parents			
Parents	I and II	III and IV	V and VI	Total
I and II	0	2	8	10
III and IV	3	2	7	12
V and VI	4	3	17	24
Total	7	7	32	46

Even more informative would be a breakdown on the basis of an assessment of the cultural and educational opportunities within the homes. Variation in these opportunities, as Burt's figures clearly illustrate, is more closely related to differences in intelligence test score than to variation in the material conditions of the homes. Moreover, very considerable differences in cultural conditions can arise, even within the same socio-economic group, and these differences could be expected to have a greater effect, lowering the correlation to a greater extent, the later the age of testing. It would likewise be of interest to have similar information for separated dizygotic twins.

A further point is that although the *range* of socio-economic background of the foster parents is a full one, extending from higher professional to unskilled manual, representation at the top end is very much less than at the bottom; and the same is true of the parents, to a slightly less extent. The objection that twins from a wide range of background would have yielded different results is best answered by not quoting individual cases but by obtaining correlations from an adequate and properly balanced sample; as Burt himself points out in a similar context (Burt, 1966, pp. 148–9), 'convenient' cases can be quoted by the throughgoing environmentalists and hereditarians alike. From Burt's account of how placements have often been carried out in the past, and of the possible access by educational psychologists to the records available, such a sample might well be forthcoming. Measures of the effect of heredity on differences in test score are always relative to the environmental differences within the group tested. No absolute partitioning of test variance is possible.

The table of correlations of the differences between the test results of the separated twins with the corresponding differences in the material and cultural conditions of the homes provides a welcome recognition of the importance of environmental conditions. Whether it is safe to conclude, however, that only a small portion of the former (intra-class) correlations can be attributed to environmental influences is another matter. It may be misleading, for instance, to read only the sample correlations, and to be tempted to dismiss any figure which fails to attain significance at a 5 % level as due to chance. The correlations from the differences in the cultural conditions are reproduced in Table 2, together with their 5 % confidence limits.

The correlation for the group test provides yet another demonstration of how mistaken it is to regard differences in test score as even a rough estimate of differences in 'innate general cognitive ability', unless there is reason to believe that all persons tested have had very similar cultural opportunities, as is surely very seldom the case. The same estimate from the individual test appears to be decidedly less rough; the supposed population correlation of zero is just within the 5 % confidence interval, so satisfying the committed hereditarian on the one hand and the 'culture-free' or 'fair' tester (if he still exists!) on the other. Even so the environmentalist could rejoin that a supposed population correlation near the top end of the interval (+0·49) is one equally consistent with the data. It is a little surprising that the correlation from the final assessment, which presumably is based on both the individual and group test, should be less than that from either of the separate tests. It would be helpful if Burt could indicate the relative weight attached to each of the tests, and also to any other evidence used in forming this assessment.

Table 2. *Correlations of differences in intelligence test results with differences in cultural conditions* ($n = 46$)

	Obtained correlation	5 % confidence limits
Group test	+0·43	0·19 – +0·62
Individual test	+0·26	–0·01 – +0·49
Final assessment	+0·15	–0·12 – +0·40

The conclusion that genetic constitution is a major influence on intelligence test score, whether individual test or final assessment, would seem to become inescapable if four conditions were to be met: first, if the separation of monozygotic twins has been carried out shortly after birth (and if the enquiries suggested by Burt are successful, sufficient numbers might be forthcoming to allow a reduction from the upper limit of 6 months that he used); secondly, if data only from twins separated into socially and, more especially, into culturally dissimilar homes are used; thirdly, if testing has been delayed until, say, near the end of the primary school stage; and, finally, if the correlations then obtained are at about the same level as those reported in Burt's present paper. To go further, however, and to argue that genetic constitution is a predominant influence, heavily outweighing all post-natal conditions, might still be unwarranted. Genetic and environmental factors are, as Burt reminds us, continually interacting from the start, and this interaction might well prove as potent as heredity alone. The need for further studies based on thorough assessments of the many different environmental conditions involved is evident.

REFERENCES

BURT, C. (1966). The genetic determination of differences in intelligence: a study of monozygotic twins reared together and apart. *Brit. J. Psychol.* **57**, 137–53.

HEBB, D. O. (1949). *The Organization of Behaviour*. New York: Wiley.

LEWIS, D. G. (1957). The normal distribution of intelligence: a critique. *Brit. J. Psychol.* **48**, 98–104.

(*Manuscript received* 12 *May* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 435–436
Printed in Great Britain

A NOTE ON ORDER OF RECALL IN SHORT-TERM MEMORY

By M. J. A. HOWE
Department of Psychology, University of Sheffield

Howe (1965) presented nine-consonant lists by ear to subjects at the rate of one item/sec, and then asked subjects to recall the parts consisting of items 1–3, 4–6 and 7–9 in the six possible orders. Retention was best when the three parts had to be recalled in an order which was the reverse of the order of presentation, i.e. when the order of recall of the items in a list was 7, 8, 9, 4, 5, 6, 1, 2, 3.

This finding seemed of interest in two ways. (*a*) There could be practical applications, telephone dialling for instance; and it is possible that efficiency could be improved by recalling items in an order different from that in which they were presented in other situations where lists have to be retained over short periods. (*b*) If such strategies can be useful, it is not impossible that human beings may use them in processing information and may be unaware that they do. The result was obtained, however, when subjects were told the recall order only after presentation of lists, and then part by part. It is of interest to know therefore whether the same result occurs in the more straightforward, and possibly more 'realistic' situation where recall order is specified in a single instruction given before presentation of lists. Complete recall order was specified *after* presentation in experiments reported by Posner, (1964) and by the author (summarized in Howe, 1965). Both found that at certain presentation rates, lists were better retained when later items were recalled first.

METHOD

Sixteen undergraduate subjects, tested individually, each heard twenty nine-consonant lists (using *C, D, F, H, L, N, R, S, T*) presented at one item/sec. The recall orders were (1) items 1, 2, 3, 4, 5, 6, 7, 8, 9 ('forward recall') and (2) 7, 8, 9, 4, 5, 6, 1, 2, 3 ('reverse recall'). They were used in different random orders for each subject, each order occurring five times in both the first and the second ten lists. Before presentation of a list the illumination of one of two lights told subjects the order in which subsequent recall was to be attempted, the light remaining on throughout presentation of the list. Subjects dictated their recall to the experimenter. They were encouraged to guess when uncertain, and to eliminate ambiguity about positions of items by indicating any consonant not being attempted by the word 'blank'. There were six practice lists to lessen the effects of warm-up and procedural difficulties. The intervals between lists were about 5 sec.

RESULTS

Table 1 shows the average number of consonants recalled in their correct positions. The prediction from previous results that recall of lists whole would be significantly better in the reverse than in the forward recall order condition was confirmed ($t = 2{\cdot}28$; D.F. 15; $P = 0{\cdot}025$, one tailed). Retention of the first three items in lists was better with forward recall and of the last three items in the reverse recall condition ($t = 4{\cdot}69$; D.F. 15, $P < 0{\cdot}01$ for 1, 2, 3; $t = 10{\cdot}46$; D.F. 15, $P < 0{\cdot}001$ for 7, 8, 9, both two-tailed). Retention of items 4, 5 and 6 was not significantly influenced by the recall order of the other items.

Table 1. *Average number of consonants correctly recalled as a function of recall order*

	Order of recall	
Items	Forward (1, 2, 3, 4, 5, 6, 7, 8, 9)	Reverse (7, 8, 9, 4, 5, 6, 1, 2, 3)
1, 2, 3	2 36	1·89
4, 5, 6	1·98	2·06
7, 8, 9	1·82	2·75
Whole lists	6·16	6·70

Discussion

The previous findings on the effect of order of recall when orders of presentation are not identical are confirmed under more straightforward conditions. There may therefore be practical applications to activities requiring retention over brief periods, and strategies of this kind may be used more generally in human information processing. However, any advantage for the amount recalled would naturally be worthless if the precise original sequence had to be reproduced. There is the further problem that the experiments of Posner (1964) and of Howe (1965) suggest that the present result would not occur at other rates of presentation, even with identical material and mode of presentation.

I should like to thank Prof. H. Kay for his help and encouragement. This work was carried out while the author held a S.R.C. Research Studentship.

References

Howe, M. J. A. (1965). Intra-list differences in short-term memory. *Quart. J. exp. Psychol.* 17, 388 42.

Posner, M. I. (1964). Rate of presentation and order of recall in immediate memory. *Brit. J. Psychol.* 55, 303–6.

(*Manuscript received* 21 *January* 1966)

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 437–509
Printed in Great Britain

PUBLICATIONS REVIEWED

Structure and Direction in Thinking. By D. E. BERLYNE. New York and London: Wiley. 1965. Pp. xi+378. 68*s*.

This book offers a contribution to the development of a theoretical language which will encompass the whole range of behaviour from the simplest to the most complex levels. Languages usually grow from certain roots under the influence of cultural exchanges. The roots of this language are firmly embedded in Hullian theory and the main acknowledged influences come from Geneva, from Russia and from work on computer thought simulation. One of the primary aims is to achieve, 'A confrontation of theoretical ideas and empirical data supplied by Piaget with those supplied by the English-language tradition'. Few people can be as well qualified to make such a confrontation as the author, whose valuable contributions to the analysis of behaviour on the one hand and translations and expositions of Piaget on the other are well known. As he received his doctorate from Yale, it is no surprise that his native psychological language includes S–R terminology. The Hullian language is intended to assimilate that of Piaget and several writers on thinking, but suffers inevitable accommodation in the process. Adaptation is evident in some fresh contributions to the analysis of thinking. The author warns that the book is not for those who do not sympathize with its primary aim of developing a theoretical language. Interest may depend also on readers' sympathy with one of the main source languages. It may be stimulating for a Hullian to see how the associationist system can be expanded.

Those who have wrestled with Piaget's own accounts of his theories will find this interpretation of great interest. It would be a pity if lack of sympathy for either source language diminished appreciation of the new things the author has to say. This book is clearly not for the beginner.

The book begins with an exposition of a 'neo-associationism' intended to embrace thinking as well as simpler forms of activity. It involves mediated responses and stimuli, and wide-ranging stimulus and response equivalences. Instead of 'rules', 'strategies', or 'plans', the author prefers Hull's concept of 'S–R generalization'. Experiments cited by others as contrary to the associationist position are offered as evidence of simultaneous generalization of stimuli and responses. Here, as often throughout the book, Hull's views are stoutly defended. The exposition of Piaget includes a clear summary of the theory relating to classes, relations, groupings and equilibrium in intellectual structures. The essential findings of the experiments on quantitative invariance are reviewed, together with the work, hitherto unavailable in English, on the effects of special training procedures on the attainment of responses belonging to the stage of concrete operations. The author argues that the demonstrations concerning quantitative invariance offer a paradigm for directed thinking preferable to the typical 'concept attainment' experiment. One contribution to the translation of Piaget's language into that of Hull is given in the suggestion that, 'If we therefore say that Piaget's operations are implicit responses organized in habit-family hierarchies, we have apparently captured the fundamentals of his conception'. The notion of habit-family hierarchy is greatly enriched by the suggestion that such hierarchies might be organized in systems with properties approximating mathematical group structure. Another happy idea illustrated on the dust cover is that compound habit-family hierarchies can be thought of as consisting of tree structures whose bases are joined together.

In order to deal with the long-standing problem of the direction and control of thinking (why this particular association at this particular juncture?) it is suggested that a new kind of response is required, one which will ensure that only legitimate steps are taken between stimulus situations. The receptor-adjusting responses of Sechenov and the operational responses of Piaget are considered to be varieties of transformational response. Such responses are thought to be analogous to the operators of Newell, Shaw and Simons's 'General Problem Solver', and to be responsible for, among other things, the schematic determination of posture, the constructive character of recall and the eduction of correlates and relations. It could be argued that a special kind of response is not required here, for surely it is a feature of most overt responses that they transform stimulus situations in reliable ways and, if implicit responses resemble overt ones, why not in this respect also? The concept of transformational response is a stimulating contribution to the analysis of behaviour and deserves careful consideration.

Three chapters discussing the motivation and dynamics of thinking extend the author's previously published theories of the role of conflict in the motivation of behaviour to include conceptual conflict. Conflict is seen as an undercurrent of all activity almost synonymous with wakefulness. A problem is described as, 'A conflict situation that gives rise to epistemic curiosity'. A final chapter on productive thinking discusses creativity and insight. The latter is largely a dialogue between Kohler and Hull, with comments by Pavlov, but no word from Harlow.

To construct a single language which will unify several others, it is necessary to show how the latter can be translated into the former. When more or less familiar terms are being replaced by more or less unfamiliar ones, depending on prior allegiances, there are bound to be some less happy renderings. It is perhaps inevitable, when the interrelations of old and new concepts have to be suggested, that unknowns should figure in the explanations of other unknowns and that the argument may sometimes seem circular. As long as psychologists continue as at the Tower of Babel, attempts to unify several discourses will be hazardous but of great potential value. This book makes a significant step towards the synthesis of psychological languages and it is worth the attention of all interested in this aim.

MARIAN ANNETT

New Directions in Psychology. II. By F. BARRON, W. C. DEMENT, W. EDWARDS *et al.*, and J. and M. OLDS. New York: Holt, Rinehart and Winston. 1965. Pp. x + 422. 28*s*.

A volume called *New Directions in Psychology* with a foreword by Prof. Theodore Newcomb appeared in 1962. It consisted of four contributions describing developments in four separate fields of psychology. The book must have been successful enough to encourage the publishers to produce three years later *New Directions in Psychology, II*, comprising four more such contributions. According to the foreword, the aim of the present volume is, as before, to report on developments in selected fields of psychology in a manner which is 'not too technical'. In fact, the contributions, although clearly presented, are on the whole not much less technical than any specialist psychological monographs.

'The Psychology of Creativity' by Frank Barron falls into two parts: (1) the measurement of creativity, and (2) the study of highly creative persons. The former describes mainly how tests of aesthetic discrimination and independence of judgement differentiate creative from non-creative individuals. The second part is a progress report of studies of highly creative groups of architects, writers, mathematicians and others. Much attention is properly given to the criteria of creativity. The relationship between creativity and intelligence, as traditionally assessed, is considered in the light of the now available evidence. Barron's thorough review ends with a section on the nurturing of creativity which, although of necessity inconclusive, is interesting and thoughtful.

'An Essay on Dreams' by William C. Dement begins with a description of physiological variations during sleep; the EEG and oculomotor activity are particularly fully and clearly described with a great deal of illustrative material. The author then examines the proposition that dreaming occurs only during REM (rapid eye movements) sleep, and concludes that such a view cannot be upheld; he hypothesizes, however, that, as far as the C.N.S. is concerned, REM sleep is like wakeful mental activity associated with sensory input from the environment, whereas non-REM sleep corresponds to other forms of conscious mental activity, viz. 'abstract thought, apperception, decision making, imagery, problem solving and so forth'. Dement has himself done much research in the field of dreaming. He reports at length on studies concerned with the content of dreams, the recall of dreams, dream deprivation, and others. All in all, this is a very valuable survey.

'Emerging Technologies for Making Decisions' by Ward Edwards, Harold Lindman and Lawrence D. Phillips starts off, as a preliminary to the topic of decision making, with the problem of the characteristics of rationality. The paper considers next a number of conceptions of probability. The technical section then deals with Bayes's theorem, which is said to be 'an elementary consequence of the fact that probabilities add up to one'. It is noteworthy that the Bayesian analysis entails the conclusion that it is proper 'to collect data until you have enough to prove your point', which runs counter to the view of orthodox statistics. Further sections on probability are concerned with empirical data, e.g. subjective judgements of relative frequencies. The empirical problem of 'real' pay-off matrices is raised, but it is pointed out that little research has as yet been done in this area. The shortness of the bibliography shows the embryonic state of the empirical study of decision making. The material in this paper is extremely clearly presented.

'Drives, Rewards and the Brain' by James and Marianne Olds is the last of the four contributions. The authors put forward the view that there is a fairly sharp differentiation between mid-brain functions and fore-brain functions. The former have to do with involuntary, reflex behaviour and the latter with voluntary, operant behaviour. Aversive behaviour appears to be predominantly controlled by the mid-brain, whereas appetitive behaviour would seem to be largely controlled by the fore-brain. Having sketched the picture, the authors proceed to adduce evidence for their views, mainly from recent experimental studies of the behavioural effects of electric stimulation of the brain, of drug effects, and of effects of partial decortication of animals. The discussion of brain functions is clear, but that of the character of drives seems somewhat recondite. Theorizing about the neuro-physiological basis of operant behaviour is highly speculative, as any such theorizing in the present state of limited factual knowledge is bound to be. All this does not detract from the value of the paper. Indeed, this and the other three papers make a very useful book.

W. SLUCKIN

Frontiers in Psychology. Edited by R. W. RUSSELL. Chicago: Scott, Foresman. 1964. Pp. 209. 20*s*.

This little book covers a wide field. It starts with philosophy and ends with Colonel Glenn's account of manned orbital space flight. Within this vast area there are, as is to be expected, some excellent pickings. The editor has watched the rapid expansion of science and technology in general and has caught the excitement of psychology's own advances and their application to present-day problems.

Seven quite separate areas are discussed: science and philosophy; hereditary potential in different environments; the application of programmed learning and teaching machines; recent research on motivation; chemistry and behaviour; intelligence, creativity and learning; and human factors in manned space flight. Each of these topics is represented by four recent papers. Of course, any such selection must be an impertinence, but it gives a flavour of psychological thinking in these fields. Most of the papers are reprints and many such as Festinger's 'Cognitive Dissonance' or Harlow's 'Mice, Monkeys, Men and Motives' are well known. It is useful to have the section on chemistry and behaviour—Prof. Russell's own interests—and refreshing to have an article on behaviour genetics in the section on hereditary potential.

I enjoyed the papers on manned space flight; Glenn's article is good reading and often good psychology. But Newell's article seems to have got in by a printer's error. It is a beautifully clear account of the Van Allen belts, the ionosphere, radiation hazards, etc., but it can only be a frontier which the psychologist is facing with every other scientist.

H. KAY

The Nature of Human Conflict. Edited by E. B. MCNEIL. New Jersey: Prentice Hall. 1965. Pp. xvi + 315. 56*s*.

This is an introduction to research on international conflict. It might serve well as a springboard for one of these interdepartmental postgraduate-and-staff seminars on international conflict which are now to be found in every serious British university. It offers an enormous number of insights from all the behavioural and social sciences, and almost every page opens up vistas of further work to be done.

But it is not what its publishers claim—a 'systematic interdisciplinary study of human conflict'. It is a collection of chapters, mostly about international conflicts, by the following authors, all Americans and all but one academics: Chadwick Alger (political science), Robert Angell (sociology), Kenneth Boulding (economics), Richard Falk (law), Ole Holsti (sociology), Daniel Katz (psychology), Donald Keys, Charles McClelland (political science), Elton McNeil (psychology), Margaret Mead (anthropology), Rhoda Metraux (anthropology), Robert North (political science), Anatol Rapoport (mathematics), David Singer (political science), Ross Stagner (psychology), Stephen Withey (psychology).

The chapter headings form a neatly methodical sequence; the chapters themselves do not. Most of the contributors seem genuinely anxious to take advantage of the experience of disciplines other than their own, and there are several exhortations, to whom it may concern, to produce a new kind of social scientist who could do interdisciplinary work. Perhaps that would have been made easier if the contributors themselves had worked together to produce a report on particular problems of management in the international system, instead of writing separate chapters, excellent though some of them are. As it is, the book contains no substantial discussion of the particular opportunities and difficulties of collaborative problem-oriented work.

Most of the chapters are based on named academic fields, e.g., the psychology of human conflict; system theory and human conflict. Some of them summarize work done, some offer conceptual frameworks, one is mainly a case study. As they average 10,000 words each, comprehensiveness is not to be expected. The baldness of some sections is disconcerting, and perhaps misleading, but this is not a reference book. (There is no index.) There is quite enough in it to lead the sceptical but concerned reader on to other books, and to continuing collaboration with other kinds of behavioural and social scientists. If that does happen, it will justify the price of the book; at $9.00 it is rather expensive, but not nearly as expensive, even to an individual buyer, as the more usual efforts to deal with international conflict. DAVID BARKLA

Conflict, Decision and Dissonance. By LEON FESTINGER. London: Tavistock Publications. 1964. Pp. vii + 163. 37*s.* 6*d.*

This latest product of Festinger's cognitive-dissonance workshop maintains, and at times surpasses, the excellence of its predecessors. It reports and comments on ten experiments concerned with the elaboration and further exploration of dissonance theory, carried out by Festinger and his students. Each experiment is presented in sufficient detail, always preceded by a theoretical discussion to show what kind of thinking led to the design of this particular study, and followed by a commentary to show us where the findings have confirmed expectations or brought us further, and where they have failed to do so or produced unexpected results. Though the material is often difficult, the style is always lucid, the presentation thoughtful and self-critical, and the book presents an integrated whole. Ingenuity in designing and bringing about experimental conditions which are crucial to the theory is matched by willingness to concede alternative explanations of the findings—and by the design of further experiments to act as tie-deciders.

Needless to say this book is no more than a progress report; it explores certain aspects of the decision-making process with the aid of dissonance theory and ties up certain loose ends, but it is in no way intended as final. It shows the need to make fine distinctions, for instance between conflict and dissonance, and between pre-decision and post-decision behaviour—and even between decisions without consequences or commitment, and decisions which definitely produce changes for the individual. It shows that before the decision is made an individual acts relatively 'objectively' in assembling and ordering his preferences, and that dissonance-reducing behaviour (e.g. by increasing the undesirability of the rejected alternative) does not start until after decision and commitment have been made. Even then, there is evidence for a transitory phase of regret, in which the rejected alternative is actually regarded as *more* desirable, before dissonance-resolving processes set in. These latter consist not so much of distorting or biasing previous information, as of a process of *re-evaluation*, and of selective exposure to new information. The nature of this selective process is illuminated by further experiments.

Much remains to be done, and there is some room for criticism. The whole trend of these experiments is to regard decision making as a *cognitive* process, whereas decisions in real life so often seem attended by frustration, impulsiveness, emotional, unconscious and irrational processes. Festinger is not unaware of these, but so far they have not been brought into his experiments. It is also hard to agree with the conclusion that dissonance-reducing phenomena do not start until after the decision has been made; sometimes, for instance when the subject is threatened by a decisional impasse, only a systematic spreading apart of the alternatives can make a decision possible. Then there is the problem of the not-infrequent decision *inversions* in several experiments, when subjects chose alternatives to which they had previously given a lower preference score. Some of the experimental tasks seem extremely unrealistic, e.g. asking

13- to 16-year-old schoolgirls to choose a vice-president for a firm which they 'owned and controlled'. If any use has been made of the subjects' introspections on their own decision-making processes, it is not reported here.

If there are gaps, that leaves us something to look forward to. In many ways the whole of our scientific endeavour is a striving for consonance and the reduction of dissonance. This book contains much to show how this can be achieved.

A. N. OPPENHEIM

Imitation of Gestures. By J. Berges and I. Lezine. London: The Spastics Society Medical Educational and Information Unit in association with Heineman. 1965. Pp. 116. 21*s*.

The development of motor skills during childhood has received relatively little attention, not because it lacks importance but because of the obvious difficulties involved in testing the younger age-groups. It is much to the authors' credit that they were able to put into practice a detailed and consistent examination of motor skills of a substantial group of normal French children aged three to six. The method used was that of 'imitation of gestures', the child being asked to copy the examiner's movements or more precisely to adopt postures identical with those assumed by him. In so far as a 'gesture' acts as a vehicle for communication, use of the word in the present context is misleading; the gestures used were not ones which primarily convey information. 'The Imitation of Hand and Arm Postures' would have been a more closely descriptive title.

The test procedures are clearly described and well illustrated, and for each age-group percentage success on each of the test items, which varied in complexity, is stated. The examiner faced the child during testing, giving scope for reversal errors, and these are separately scored and discussed. Results are also presented on a number of children born prematurely, which show the expected relative deficits in performance.

With the exception of a preface by Ajuriaguerra, the text is admirably free from theorizing, and the work is presented as an empirical study. The findings will be of interest to workers concerned with human motor development.

The work runs into difficulties only when, as the authors propose, it is regarded as providing a standard test for use by clinicians. It is quite unclear how results relate to motor development or to general intelligence; the fact that the 'test' correlates 0·33 with the Goodenough Draw-a-man test and 0·494 with the Grace Arthur mannequin puzzle does not really help.

There is a further section on the ability to name body parts (somewhat coyly selected) the findings of which can presumably not automatically be extended to English-speaking children.

In sum, the work here reviewed provides carefully collected data on the ability of children aged three to six to position their hands and arms in certain ways. This is of research interest, but there is as yet no good reason why clinicians should feel obliged to perform the test on their patients. The danger arises that deficits on this test may be attributed to faults in that polymorphic entity the body image, and that such attribution be regarded as meaningful.

The translation is excellent and the format reaches the high standard customary in publications associated with the Spastics Society.

M. KINSBOURNE

Leaders, Groups and Influence. By E. P. Hollander. New York: Oxford University Press. 1964. Pp. xi + 256. 40*s*.

The title of this book suggests that it presents a systematic treatment of the relation between leaders or leadership, group structure and function and the general process of social influence. And indeed psychologists acquainted with Dr Hollander's work might well look forward to precisely such a contribution from him. It is, unfortunately, not contained in the present volume.

In essence this is an anthology of Dr Hollander's papers. Seventeen of the twenty chapters are reprints or adaptations of articles already published in various journals. They are preceded by a brief historical introduction and rounded off by a short chapter suggesting some of the implications of various concepts and findings for further work. The remaining new chapter is a somewhat rambling discussion of organizational leadership.

The basic character of the book is thus that of an anthology. In principle, there is some value in collecting a series of papers on related subjects and combining them into one volume. At the very least it is convenient for the reader and it may offer an opportunity to trace the development of an author's ideas. Dr Hollander's book does offer these advantages up to a point: in editing his papers he has at times nicely juxtaposed articles concerned with the elaboration of ideas and experimental work demonstrating their usefulness. Yet even so the book does not make a coherent whole. For instance, nearly one-third of it is devoted to expounding the peer-nomination technique, its nature, validity and reliability; this makes for a rather disproportionate emphasis on methodology and even the usefulness of this is limited by the absence of any systematic assessment of this technique compared with others.

Much the most valuable section of the book is contained in the last seven chapters. There Dr Hollander sets out his dynamic view of conformity and argues for his concept of 'idiosyncrasy credit'. He holds that conformity consists of an individual behaving as *he* is expected to behave—in contrast to conformity considered as behaviour which is judged 'correct' by some general or abstract criteria. Thus the individual who conforms to the expectancies he himself has built up in others accumulates 'idiosyncrasy credits' which may ultimately allow him to deviate and to become a leader and innovator. This scheme is then used to resolve the paradox that leaders attain high status by conforming, while yet being expected to innovate once they have attained leadership; for in essence, when conformity is seen to reside in the consistency of the behaviour of the individual, the potential leader is, in a sense, always true to himself. All he has to do is to obtain 'credit' for his own form of behaviour within the group.

But, although this final part of the book is particularly stimulating, it also leaves the reader dissatisfied. There is no systematic treatment of the different forms and functions of leadership, no discussion of the effect of the nature or function of the group. Indeed, although the word 'groups' features in the title of the book, the reader will know little more about group behaviour when he has finished reading it.

The basic weakness is simply this: papers prepared for journals are almost inevitably limited in their scope; they do not usually allow a full examination of a set of problems and its implications. Similarly, a mere anthology of papers leaves the reader dissatisfied because of all that has been left unsaid. This is the case with the present volume. There is much of interest here, but none of it is truly worked out. The main benefit is the advantage of having a variety of papers in one volume. One can only hope that it will not be long before Dr Hollander elaborates on this work to present us with a systematic treatment of leadership, groups and influence. All that this book does is to suggest that he would then make a most valuable contribution.

R. P. KELVIN

Attitude and Attitude Change. By CAROLYN W. SHERIF, MUZAFER SHERIF and ROGER E. NEBERGALL. Philadelphia and London: W. B. Saunders. 1965. Pp. 264. 58*s*.

In the kaleidoscopically changing fashions of American social psychology, the work of Muzafer and Carolyn Sherif stands out for consistency over many years in its unwavering adherence to a few basic assumptions. First, they rely on 'a disciplined phenomenology as a basis for theorizing'. Secondly, they are concerned with a person's involvement in an issue, postulating that it is the task of social psychology to study attitudes which really matter to people, and not carefully sterilized forms of behaviour deprived of meaning. Thirdly, they maintain that attitude and attitude change must be viewed as judgement processes, akin to psychophysical judgements, in which evaluative, affective and behavioural components are combined in the manner a person categorizes a communication which is different from his own approach, by using the latter as an anchor. Fourthly, they regard the social context of behaviour, particularly an individual's reference groups, as a basic factor in the development of his attitudes and in attitude change.

While this book contains several previously unpublished studies mostly conducted during the 1960 American election campaign, its major value lies in presenting in a coherent form all the concepts and methods which the authors have used here and elsewhere in the study of attitude and attitude change.

Central to the Sherifs' approach is the concept of latitude of acceptance and latitude of rejection which must be determined in order to make predictions about an attitudinal position possible. The Sherifs operate with attitude scales ordered in a continuum from one logical extreme to its opposite. Individuals whose attitudes are to be assessed first choose the statement which comes closest to their own position and the statement which they find most objectionable; then they indicate the range of tolerance (latitude of acceptance) and their range of rejection (latitude of rejection) for other positions in the scale. The remaining items, if any, are the area of non-commitment. The latitude of rejection on a particular issue emerges as a central measure of ego involvement, which is crucial for making predictions about whether a discrepant communication will be dealt with by assimilation or contrast.

The exposition of these ideas and their confrontation with other approaches—Guttman scaling and dissonance theory receive some hard-hitting criticisms—make good psychological sense; the research data are, of course, competently presented. And yet altogether this book reads more like a promise for the future than a definitive statement. In part this is due to the authors' deliberate caution in describing their work as an approach rather than as a theory or model; in part, however, it is due to the nature of the research results, which are boring. All they do is to support the approach, while one expects from the Sherifs that they will use an approach in the solution of a problem. Perhaps this is too much to expect at this stage of social psychology; or perhaps the Sherifs' next book will show that it can be done.

MARIE JAHODA

Spatial Ability. Its Educational and Social Significance. By I. MACFARLANE SMITH. London: University of London Press. 1964. Pp. 408. 35*s*.

For this reviewer the chief question raised by Dr Macfarlane Smith's interesting book is: what *is* spatial ability, and how far can one discuss it as a unitary phenomenon, or equate it with, say, the '*k*' factor? The title perhaps suggests more general implications than are dealt with, although this does not lessen the usefulness of this book; it is not that the author confines himself to a narrow viewpoint, but that the subject is a very wide one.

Reading a map or blueprint, painting a picture, designing a house—all these presumably require spatial ability. One can argue further that dressmakers, surgeons, dancers and successful trapeze artistes require in their own ways as much spatial ability as engineers, sculptors and architects. If the manipulation of the body in space reflects spatial ability as much as the bodily manipulation of three-dimensional objects, and of two-dimensional representational space, then it is necessary to discover how these various aspects are related, through study of their genesis and development.

Dr Macfarlane Smith does not set out to deal with spatial ability in this developmental, Piagetian sense (in fact Piaget is not mentioned at all). Roughly the first half of his book is concerned to make a case—and the arguments are detailed and forceful—for the use of pencil and paper 'spatial' tests in the selection of young people who wish to specialize in technical, scientific and mathematical subjects in secondary and further education. Where success at this level is concerned, spatial tests are useful as predictors and have been grossly neglected in the past.

Are adults in occupations considered to require high spatial ability also differentiated from others by these tests? Perhaps it is unfair to wish for yet more direct evidence on this point, but it is not clear how far American findings support the tests, and, perhaps because, as Dr Macfarlane Smith suggests, the better spatial tests were not always used in the Services in this country, evidence here is not clear either. He quotes Vernon and Parry (1949) as saying 'that while there is ample evidence of the vocational value of spatial tests among adolescents with little mechanical experience, very rarely were they found to assist in the selection of adult mechanics in the Services'. Can we equate the '*k*' factor with spatial ability in 'spatially sophisticated' adults?

The author's approach is primarily that of the psychometrist, and the chapter on the spatial factor and its subdivisions will be invaluable to the worker in this field, for it is an authoritative account of research and controversy.

In the latter part of his book Dr Macfarlane Smith puts forward some challenging ideas, including the suggestion that spatial tests may be true abstractions tests, and possibly 'super-*g*-tests'. He reports interesting research on the spatial difficulties of cerebral-palsied and brain-injured children, and following leucotomy and brain injury in adults. His speculations about the relationship between spatial ability and temperament are provocative also. Compare his comment based on research by Terman and Miles (1936) that 'members of a profession which tends to recruit spatially gifted individuals will tend to show high mean scores on the 'M/F questionnaire' (i.e. high masculinity) with MacKinnon's finding that all three groups of architects studied by IPAR had an extremely high peak on the femininity scale—the more creative the architects the higher the femininity.

All in all, this is a book deserving of attention. Some psychologists may raise an eyebrow at a few of the author's arguments, but Dr Macfarlane Smith will be justified if further research is prompted thereby.

M. O. TYSON

Affect, Cognition and Personality. Edited by S. TOMKINS AND C. IZARD. New York: Springer. 1965. Pp. xii+464. $10.00.

This book is 'an expanded version' of a symposium on affect held at the 1964 meeting of the A.P.A. and contains fourteen papers contributed by twenty-one psychologists including Gardner Murphy, Ralph Exline and the editors themselves. The subject-matter of the papers varies widely and the book suffers from an absence of direction and integration.

With a few exceptions the quality of the contributions is disappointingly poor. The more interesting papers are found in sections 6 and 7 and deal with 'affect and facial responses' and with methods of measuring affect. Leventhal and Sharp describe a study of the facial expressions of people undergoing genuine stress (childbirth) and their findings suggest that consistent changes do occur and that they can be reliably assessed. Exline's work on 'mutual glances' and affective relations is carefully presented and contains some interesting observations. For example, he cautiously offers the following generalization: 'Women...seem to seek out the eyes of those they like, or to whom they are momentarily attracted, whereas men do not attempt to increase their contact with the preferred so much as to avoid contact with the less preferred'.

Nowlis gives a full account of the current status of the Mood Adjective Check List. This relatively crude test can provide a surprisingly effective measure in a range of situations but exclusive reliance on the MACL should be avoided. Ekman's paper on communication through non-verbal behaviour is well organized; the argument is sustained throughout and is used to impose order on a number of sequential experiments. His systematic exploration of the topic is a model which the other contributors (and the editors) would have done well to imitate. Like the work on facial expressions mentioned above, Ekman's research is pertinent to Gellhorn's views on proprioception and emotional reactions.

In addition to editing the book, Tomkins provides two contributions of his own. The first is overshadowed by some loose journalistic sociology and a nebulous plea for commitment. This tendentious and ethnocentric piece is a good example of the dangers inherent in Tomkins's idea of commitment. His discovery that the Rorschach test is 'an essentially left wing method' is (unintentionally) amusing.

Another passage which provides a pleasant diversion occurs in a paper by Izard *et al.* Commenting on a result obtained in an experiment on the effects of affective picture stimuli they state: 'We never doubted that the pictures had some sex stimulus value, but we hoped this would be overshadowed by the positive affect aroused.' Regrettably, this drawn-out collection contains few such light touches and many dreary sections. Some of the papers are virtually sterile.

One must conclude, unfortunately, that this patchwork of papers does little to advance the study of affect. It seems unlikely to provoke much interest.

S. RACHMAN

Emotions and Emotional Disorder. A Neurophysiological Study. By ERNST GELLHORN and G. N. LOOFBOURROW. New York, Evanston and London: Hoeber Medical Division, Harper and Row. 1963. Pp. xii+496. 90*s*.

The senior author of the book has, for many years, and with considerable distinction, devoted himself to the neurophysiology of the hypothalamus. In recent years others have enthusiastically pursued less circumscribed areas of the brain-stem, publishing papers paralleling those of Gellhorn, who has continued to write in terms of the hypothalamus alone. It must be said that this book conveyed an unnecessary impression of a sacred hypothalamus and of slightly injured defenders striving to save for mankind control over the hypothalamus, a control the importance of which, 'with all its personal, social and political implications, can hardly be exaggerated'.

The book opens with basic anatomy and neurophysiology. The evolution of the limbic structures, the theory of Papez, the work of Bard, of Olds and of Kluver and Bucy are summarized but not developed. Gellhorn introduces his concept of 'tuning' in the autonomic nervous system: that the hypothalamus can undergo physiological 'tuning' so that it can be over-ready to cause sympathetic response on the one hand or parasympathetic response on the other. Much space is devoted to blood-pressure control. He goes on to espouse the mecholyl test, giving references to his own work but not to that of others who have failed to find the test valid or useful. He progresses to sleep and wakefulness, conditioned reflexes, experimental neuroses, stress and the adrenal cortex, vascular hypertension, peptic ulcer, nausea, fear, the psychoses, shock therapy and theories of emotion. In every instance he keeps our attention fixed upon the hypothalamus.

Gellhorn's sympathies are for the German psychiatry of an older era. Emotional disorders may be cured by 'removal of focal infection' but otherwise we must look to the hypothalamus. It is not that learning should be disregarded, for 'the hypothalamus plays a dominant role in the establishment of conditioned reflexes'. Evidently, however, learning plays no part in anxiety states. Anxiety, for Gellhorn, is a component of 'sympathetic attacks' due to an unfortunate 'hypothalamic balance'. Schizophrenia and mongolism are introduced with hints of hypothalamic causation. More boldly comes the assertion 'hypothalamic–cortical discharges...account for, or contribute to, the mental changes seen in endogenous depressions and other functional psychoses'. 'By shifting the hypothalamic balance' to the parasympathetic side 'we produce depressions' or, in the opposite direction, mania, which 'explains the beneficial effect of electroshock therapy'. One could have wished for some evidence. He falls into the trap of equating human depression with inertia when, in fact, agitation is a commoner *accompaniment* of depression, while the essence of mania is not the overactivity but the mood of elation and gaiety.

Anorexia nervosa, for Gellhorn, is a 'neurological disorder' which is 'often associated with depression'. We are told in italics that '*emotions and instinctual drives occur only in the alerted organism*'. One does not have to be a Freudian to have nightmares and wet dreams while asleep.

Barbiturates and lysergic acid, for Gellhorn, have their principal actions upon the cerebral cortex, reserpine and chlorpromazine upon the hypothalamus. These statements exemplify the whole of the writing—the sites of action he favours are unquestionably important. No one would dispute that—but that does not justify so little attention to the rest of the brain, the biased choice of references (not always dealing with the contiguous topic) nor endless statements of faith presented as fact.

The psychology is naïve and at times obscure. On the subject of attention—'reduction of the sensory input to the cerebral cortex which makes concentration on a certain object possible' thanks to the action of the hypothalamus which makes possible activation of a much larger area in the cortex and so produces recognition.

I cannot recommend this book. The quantity of factual information contained, and some worth-while commentary, lose their value in a setting that lacks impartiality and humility. A book for the critical, advanced student only.

IAN OSWALD

Visual Capabilities in the Space Environment. Edited by C. A. Baker. Oxford: Pergamon Press. 1965. Pp. vii + 203. 63*s*.

This volume is a reprint of the thirteen papers in the June 1963 issue of *Human Factors*, with three additional papers. Most aspects of vision of importance in determining man's role and limitations in space are included. There are, therefore, papers on visual search tasks, distance estimation, and vigilance with special reference to discrepancies between laboratory and operational findings. The visual problems in space rendezvous, extended space flight, and moon landings are described, and man's ability to perform survey and reconnaissance tasks in space assessed. It is suggested that visual performance will be affected by various environmental factors, such as atmospheric contaminants within the space vehicle. The effects on vision of gravitational forces and vibration are discussed with emphasis on possible methods of amelioration. The interaction between visual and other sources of information can produce conflicting evidence, and the nature and consequences of such conflicts are predicted. The choice of information displayed to the astronaut depends on the space environment, the total system, and the purpose of the mission. The quality and quantity of information must match the mission requirements. Methods have been devised for displaying new forms of information, for example to permit control of the space vehicle. Visual problems encountered in attempts to simulate the space environment are noted.

Progress in this area is so rapid that any book is out of date in some respects when it appears. Nevertheless, this volume succeeds in surveying the main areas of interest, describing the problems, recording progress, and showing how many of the problems interact. All this is worth while, and on the whole justifies this reprint.

Inevitably the quality of the writing and of the information is uneven. This is not just because there are the usual differences between authors in their approach, but also because the quality and detail of the available information varies greatly, ranging from fact to speculation and anecdote. Much of the latter, such as astronauts' reports, is of interest in its own right without being psychologically insightful. Certain of the more theoretical papers, of greater relevance to psychology, are centred on the practical problem of man in space rather than on a theoretical issue, so that the theories and models used can be judged by their practical value. Perhaps of most interest to the psychologist are the papers which remain relevant to the main theme while adopting a broad approach. Although it is invidious to single out one paper, the survey of vigilance by Jerison and Pickett deserves special recommendation.

This volume is a good source of references on most of the topics included. It is a useful record of the development of a rapidly expanding area and is of considerable general interest. Its most psychological papers are usually adequate, though seldom inspired. It has an index which is one of the most useless and perfunctory that I have ever seen. And what a pity that this volume, on a subject which lends itself so well to illustration, should have only five photographs, none of high quality.

V. DAVID HOPKIN

Mach Bands: Quantitative Studies on Neural Networks in the Retina. By F. RATLIFF. San Francisco: Holden-Day. 1965. Pp. xii + 365. $15.25.

Dr Ratliff is widely known as a student of retinal physiology. His interests extend to the psychology of visual perception and the theory of knowledge, and all these have contributed to his interest in Ernst Mach and to the present book. Mach published his first account of the accentuation of brightness contrast at visual boundaries a hundred years ago, though these papers were largely overlooked, and not the least valuable feature of the book, and a good point at which to start, are the translations of Mach's six important papers on this subject, which are included as its second part. His simple yet ingenious experiments, and precise deductions from them, led him to an account of the Mach bands essentially the same as that we now hold. His belief that 'the excitations of two [retinal] points reciprocally obstruct their discharge to the sensorium' anticipated present views, and he offered a mathematical description of these inhibitory effects.

The book begins with an account of Mach's life, work and philosophy; it traces the development of his theory of knowledge, perhaps the most influential sector of his thought, contributing as it did to the development of relativity theory. This is followed by a review of psychophysical work on the Mach bands, especially attempts to measure them and the effects of eye movements and stabilized vision. Next comes a critical account of six mathematical models of the retina which are capable of explaining the accentuation of contrast at contours, including Mach's, the earliest, and later ones due to Huggins and Licklider, von Békésy, Hartline and Ratliff, and others. They are reduced to a common notation and it is shown that they are basically similar.

Physiological evidence for retinal inhibitory interaction and the structure of receptive fields is discussed, and a concluding chapter deals with 'appearance and reality'. This contains a number of historical anecdotes in which properties of the observer—such as the production of Mach bands—were confused with properties of the object observed. These are of interest in their own right, but nothing which is said about them would be novel to a philosopher.

This book should be taken not as an act of piety to Mach, nor as the coherent development of a single theme, but as a set of essays or reviews on overlapping topics. Inevitably there are loose ends, but it is a volume which all workers in this area will find of value, and for which we should be grateful.

MICHEL TREISMAN

Genetics and the Social Behaviour of the Dog. By J. P. Scott and J. L. Fuller. Chicago and London: University of Chicago Press. 1965. Pp. xviii + 468. 90*s.*

Canine Behaviour. By M. W. Fox. Springfield, Ill.: Thomas. 1965. Pp. xiii + 137. $6.75.

All first-year undergraduate psychology students are aware of the nature/nurture controversy and its chief proponents in the form of Watson and McDougall. Few however of the introductory texts carry the issue into more recent history. During the 1930's and 1940's research on the genetic basis of behaviour was largely neglected, apart that is from the somewhat unfruitful controversy about the extent to which intelligence test scores were genetically determined and consequently resistant to modification by special forms of training. The emotional revulsion amongst many scientists to the exaggerated claims of the eugenicists of the period, particularly in the light of racist doctrines propagated by certain political organizations, was hardly a spur to empirical work.

The past decade has seen an enormous increase in the amount of research on the genetic basis of behaviour and on the evolution of behaviour. At the 10th International Congress of Genetics in 1958 two papers on behaviour genetics were presented, one by Scott, the other by Fuller. By the 11th Congress five years later the two papers had grown to a full session and a symposium, in the course of which seventeen scientists reported their work.

Scott and Fuller in the van of behaviour genetics have now produced an extremely important book which may well become one of the great books of experimental psychology. The work reported in this volume is the outcome of some thirteen years of research on a project known as 'Genetics and the Social Behaviour of Mammals', carried out at the Jackson Memorial Laboratory at Bar Harbor. One of the most significant features of the whole research programme is that it represents a systematic, sophisticated and serious attempt to see psychology, particularly comparative psychology, in the framework of modern evolutionary biology. The authors are concerned with populations not typologies. The essence of the work is an analysis of the role of hereditary factors in the *development* of behaviour. They reared puppies in a carefully controlled environment, trained and tested the young of several breeds and attempted a genetic analysis.

The first of the six chapters in section 1, 'A school for dogs', outlines the investigation and what is to follow; detailed descriptions are given of the conditions under which the animals are reared and the facilities for testing them. Reading these made the reviewer wonder if comparative psychologists in this country could ever hope to get adequate facilities for large-scale experimentation. Some refreshing self criticism is also offered in this chapter. 'Looking back on the experiment we now see how we might have organized it in a better fashion' (p. 11). The second chapter discusses the probable origin of the dog and its relationship to its wild relations, the wolf, coyote and jackal. Here also is an absorbing survey of the evidence for man's role as a selective agent in shaping the evolution of the breeds used in the study—fox terriers, basenjis, sheepdogs, beagles and spaniels. One interesting piece of information is that on the basis of the available evidence it would appear that the domestication of dogs took place around 8000 B.C.

Chapter 3, 'The social behaviour of dogs and wolves', deals with the choice of the behaviour pattern as the natural unit of behaviour and with Scott's excellent work on the ethology of the dog. Scott has classified 103 behaviour patterns shown by dogs into nine categories, in terms of their adaptive significance. These have then been compared with what is known of wolves and the dogs' other wild relatives. Chapter 4 examines the major stages of development; neonatal, dominated by the process of intuition; transition, a period of rapid acquisition and development of adult behaviour patterns; socialization, the development of primary social relationships, and the juvenile phase when the skills necessary for an independent existence grow and develop.

The critical period, '...a special time in life when a small amount of experience will produce a great effect on later behaviour', is the subject of chapter 2. Here Scott and Fuller discuss the evidence for the critical period early in life when learning, particularly social attachments to members of a group or a human master, must occur if they are to develop properly. The final chapter of this first section of the book deals with the development of social relationships between mother and offspring, between litter-mates, and between the puppies and their handlers.

Following the long introductory section the next two sections are devoted to data and its analysis. In the region of 500 pure-bred and experimentally crossed puppies were investigated through the first year of their life in over thirty situations. Numerous breed differences are reported; differences in emotional reactivity, developmental differences in autonomic functioning, in training situations and on problem-solving tests. The development of physical differences and their relationship to behaviour is also treated. It is also in this section of the book that we see revealed in detail the magnitude of the task of searching for and analysing the genetic determinants of behaviour. The dog has thirty-nine pairs of chromosomes, developmental phases lasting more than a year and a very extensive behavioural repertoire.

The last section of the book deals with implications of the work for dog breeding, and here there are important practical considerations such as the training of guide dogs, and the evolution of dogs and men. The final chapter, 'Towards a science of social genetics', is in some ways the most stimulating. Ideas spark off on almost every page. The impact of the work is far reaching and it should be compulsory reading not only for those interested in comparative psychology and behaviour genetics, but for all psychologists.

Fox on dogs is a small monograph of interest to comparative psychologists, dog breeders and veterinary surgeons. The first chapter dealing with the domestication of the dog is elegantly written and illustrated with some excellent photographs of antiquities depicting dogs. Chapter 2, 28 pages and the longest in the book, is concerned with early development and the neuro-ontogeny of the dog. It is perhaps the most interesting and detailed of the chapters, and deals with critical periods of development, with the development of reflexes and with neurological maturation. This is the field in which the author himself has published many original studies. Other chapters deal with normal adult behavioural development, with the genetics of behaviour, with behavioural abnormalities and with relations between dogs and men. The chapters on various behavioural abnormalities will be of great interest to the veterinary practitioner and the dog trainer and may indeed be the precursor to a more general acceptance of animal psychiatry. The final chapter dealing with man–dog relationships is both interesting and amusing. A case of a displaced Oedipus complex in a dog is described, along with some illuminating ideas about human personality and behaviour as exemplified through the role of dog owner.

The volume should be useful to comparative psychologists and stimulating to psychiatrists as well as dog lovers of various kinds. Like all the Thomas publications, it is lavishly produced but far too costly.

KEVIN CONNOLLY

Hallucinations et Réalité Perceptive. By HENRI FAURÉ. Paris: Presses Universitaires de France. 1965. Pp. 255. F. 18.

The first two parts of this book consist of a lengthy exposition and discussion of the literature on the contents of dreams, hallucinations and delusions, especially the apparent reality of objects figuring in these and their positions in space. In hallucinations particularly, the author concludes, real objects are transformed or omitted, imaginary objects are inserted among real ones, and the whole spatial field may be re-structured in accordance with hallucinatory symbolism. Thus, there is a strange interaction between hallucination and perception which does not, for instance, exist in dreams. This conclusion might have been more clearly defined and established had the author examined and classified a number of reports of actual hallucinations; but unfortunately he has based his argument entirely on the vague generalizations of psychiatrists and of writers such as J.-P. Sartre.

The author did, however, attempt to investigate the relationships between perception, symbolic imagery and hallucination by means of a projection test he designed, called the 'Village' test. In this test, described in the third part of the book, subjects were given a number of small models of houses, trees, etc., together with more indeterminate blocks of wood, and were asked to construct a 'village'. It appeared that those mentally abnormal adults who were subject to hallucinations showed a strong tendency to introduce an unusual number of vertical features, even objects normally placed in a horizontal position. These so-called 'pylons' were infrequent with non-hallucinated adults, even those suffering from mental illness. Partly on the basis of the subjects' free associations about their constructions, a variety of symbolic meanings was attributed to the 'pylons' related to father figures, either real or archetypal. These presumably were unusually forceful in determining the responses of the hallucinated adults. But 'pylons' also occurred quite frequently in the constructions of maladjusted children and adolescents, in whom, it is suggested, infantile domination by the father was powerful enough to produce the symbolic 'pylons' even though they were not subject to hallucinations.

Though some of the author's conclusions are interesting, they could be substantiated only by a systematic study, including some measurement of the frequency of occurence of 'pylons' in the constructions of normal children and adults and of those with various types of mental disorder. What constitutes an unusual frequency of 'pylon' would also have to be clearly defined; some of the illustrations of 'pylons' given in the text are not at all convincing. Finally, the whole argument is so wordy, diffuse and difficult to follow that it can be accepted only with considerable scepticism.

M. D. VERNON

Criminology in Transition. Essays in Honour of Hermann Mannheim. Edited by TADEUSZ GRYGIER, HOWARD JONES and JOHN SPENCER. London: Tavistock Publications. 1965. Pp. xix + 308. 45*s*.

This volume of essays is a fit tribute to Dr Mannheim, who has done so much to put criminology on the academic map. It is, in fact, noteworthy that the majority of the contributors were his students and now hold academic appointments. Dr Mannheim's own contributions to criminology have covered a wide range, and it is appropriate that the essays in this volume do the same. The problems of punishment and sentencing, prison and after-care, prediction studies and crime in central Africa all receive treatment.

To the psychologist the two contributions of most interest are those by Dr Gibbens and Prof. Grygier. Dr Gibbens discusses the problem of classifying offenders in the interest of finding appropriate treatment, and also the application of the Mannheim–Wilkins prediction tables to his own Borstal research. Prof. Grygier, in a somewhat over-elaborate essay, expounds his concept of 'social progression'. His argument is that we talk too much about 'regression to a norm' and forget another process which is of paramount importance to the criminologist. It may well be that there are pressures brought to bear to bring us back to conformity if we stray, but, when we are labelled or label ourselves as in some way deviant, we tend to pursue our deviant path and become more so—instead of 'regressing', we 'progress' away from the norms, and the more society frowns upon us, the further we tend to go. This is, indeed, an important theme, as has recently been shown by Leslie Wilkins in his book, *Social Deviance*.

It is impossible to mention all the papers, and selection may seem as little invidious, but there are three papers of general as well as specifically criminological interest. They are: two on prisons and one on white-collar crime. Dr T. P. Morris writes about the sociology of the prison, surveying the literature and drawing on his own researches into the social structure of prison society. Prof. Norval Morris surveys the history of imprisonment, and looks forward to its disappearance from the scene in its present form. This was a topic in which Dr Mannheim was interested, and he was also interested in 'white-collar crime', that is to say crimes committed by 'respectable' men in the course of their business. Prof. Spencer reports on his interviews with thirty men who were imprisoned for committing such offences. But besides this he raises interesting questions about the appropriateness and efficacy of using criminal charges to control the anti-social behaviour of business men, illustrating his points by a comparative study of the way in which monopolies are handled in Great Britain, Canada and the United States. This raises the question of what behaviour a society thinks fit to label 'criminal' and what they prefer to deal with by less offensive methods.

J. W. H. SPROTT

Juvenile Delinquency: Research and Theory. Edited by H. C. QUAY. London: Van Nostrand Co. 1965. Pp. ix + 350. 55*s*.

Of the ten contributors to this book, five are professors of psychology at Illinois and elsewhere, and so the reader may expect a sophisticated exposition of what psychology has contributed, and can contribute, to the study of juvenile delinquency. However, this book is more oriented to sociology than to psychology. Perhaps the best review of interest to psychologists is the chapter by Nathan Caplan concerning the relevance of intellectual measures in the study of delinquency.

It is hardly surprising that the book has a strongly transatlantic flavour. It would be invidious to compare the literary styles of the different contributors; some of them write in a variant of the English language which may be irritating to British readers. When this reviewer was obliged to read certain American texts as a student, he developed a practice of making an English précis of the most difficult texts in order to absorb the content. Such a practice might help in certain sections of this book.

The chapter which may surprise many criminologists is that by Briggs and Wirt entitled 'Prediction'. It contains the following observation: 'The history of delinquency prediction does not include the development of any system of prediction within the context of an organization which was itself trying to solve certain administrative decisions'. In view of the history of the parole board studies in America, and the Borstal Study of Mannheim and Wilkins, such an observation does not seem to make sense. However, by 'prediction' all that these authors mean is the prediction of future delinquency in children who have not been convicted; they take no account at all of studies of recidivism. This is indeed a novel way of approaching the subject, for it entails disregarding entirely all the work with older offenders which is relevant both in content and methodology.

Criminology is an unsatisfactory subject to many people, because it embraces so many disciplines. Sociologists, psychologists, psychiatrists and statisticians are all concerned in it, and so are politicians, professional do-gooders and priests. All of them may criticize a book such as this from their own point of view. No doubt it will go on some book-lists of required reading for students, and various specialists will refer to it in their seminars with their own specialist slant. The danger is that such a book may be used as a short cut to knowledge in a cross-disciplinary sense. The general failing of this particular book is that it reports too much work with too little *critical* appraisement of the work cited. The student may regurgitate much of what he reads in ignorance of the real problems which underlie the material.

From the point of view of contemporary psychology, a lot of work goes unmentioned in this book, and a lot of questions unasked. The section that deals with 'The meaning of delinquency' gives a 'broadly psychological frame of reference' in the form of eight numbered topics. Mowrer gets a very small paragraph here, but the whole area of learning theory in relation to delinquency, which excites controversy in this country, is simply not mentioned throughout the book.

The book is well presented in the form of nine chapters, each with a separate reference list. It may prove a useful reference manual for students, if they are aware of its limitations.

H. B. GIBSON

Organizational Psychology. By E. H. SCHEIN. Englewood Cliffs, N.J.: Prentice-Hall. 1965. Pp. xi+114. 12*s*. 6*d*.

This book, fourteenth in the Foundation of Modern Psychology Series edited by R. S. Lazarus, is a recognition of the shifting emphasis in applied social psychology from the basic work group to the organization. For the researcher attracted by the possibility of combining both 'macro' and 'micro' approaches to social questions, this can be an exciting and rewarding field of study: a meeting point of anthropology, economics, psychology and sociology. But just how one psychologist views the basic question of integrating the individual and the organization is outlined in this excellent primer by Prof. Schein.

The author introduces his subject with a somewhat traditional definition of organization. It is 'the rational co-ordination of the activities of a number of people for the achievement of some common explicit purpose or goal through division of labour and function and through a hierarchy of authority and responsibility'. Within these bounds Prof. Schein analyses the psychological problems of organization. Beginning first with questions of staffing—recruitment, selection, training and allocation, he then turns to the links between individual and institution. Here he uses Etzioni's typology to describe the nature of power and authority and the basic kinds of human involvement. In the choice of methods of co-ordinating human effort, assumptions are made about behaviour. Schein identifies these as Adam Smith's 'rational–economic' man, Elton Mayo's 'social' man, Goldstein and Maslow's 'self-actualizing' man and his own 'complex' man which embraces aspects of the three preceding assumptions. A third chapter is devoted to group and intergroup relations with emphasis on the means of dealing with conflict.

At this point the author takes a new tack and deliberately moves away from the course prescribed by his earlier definition of organization. This, he argues, is necessary because 'the internal dynamics of organization generate forces which are neither predicted nor very easily integrated into such a definition'. Instead he turns to the systems conception and in so doing gives pre-eminence to the work of the Tavistock Institute of Human Relations, in particular Trist's 'socio-technical' model and Rice's 'open-systems' model. Importance is given to the study of the interactions between concern and environment. Finally, there is a chapter on organizational efficiency argued also in systems terms. Survival is seen to be the primary objective of the concern, which can best be achieved through possessing the characteristics of 'adaptability', 'sense of identity' and 'capacity to test reality', regardless of function.

It is not until the final chapter that one feels that Prof. Schein might have set out with a systems definition of organization, no matter how messy. The last chapter should really come first. It is upon an awareness of objectives and effectiveness that the concern is built, methods of co-ordination developed, with their underlying assumptions about behaviour, and the organization staffed. The true significance of historical circumstances can only be fully appreciated if the concern is viewed as an open-system. In a changing environment, the author notes, the essential human problems of the organization are to recruit and develop personnel with the kinds of flexibility and adaptability that are needed for survival. In this context, it becomes feasible to identify the assumptions which the concern should make about human behaviour. There is evidence in plenty to suggest that the prevailing methods of control in organization are based upon assumptions of economic and social man and produce submissiveness and dependence among employees. These are human characteristics associated with ineffective performance in change. By tradition, our social institutions have proved as restrictive of the individual's pursuit of growth as they have proved helpful in meeting his deficiencies. The environment of the present day calls for organizations with systems for co-ordinating effort based upon assumptions of self-actualizing man.

Prof. Schein's book is also noteworthy for the clarity of its style and language, the judicious use of research material, its chapter summaries and the printing in italics of critical points. All in all, it is one of the best introductory texts the reviewer has read.

DENNIS PYM

Personality and Personal Illness. By G. A. FOULDS. London: Tavistock Publications. 1965. Pp. xi + 344. 55*s*.

Clinical psychologists often seem ferociously ambivalent in their attitude to psychiatry in general and to diagnostic psychiatry in particular. They publish gleeful accounts of the inadequacies and illogicalities of psychiatric classification, while utilizing it as criterion for their most treasured tests. They often maintain an impressive ignorance of psychiatric teaching on classification, while earning their daily bread by slotting patients into diagnostic categories. Foulds's book is thus unusual in that it represents an intelligent attempt to integrate psychiatric teaching and observations with psychological (and logical) methodology in order to produce a rational and useful classification system for personal illness—Foulds's interesting replacement for the older soul–body compound 'mental illness'.

Foulds's central thesis is that personal illness lies along a continuum defined by an increasing failure to establish or maintain mutual personal relationships. The stages within the continuum denote the point at which various psychiatric groups have come to rest, but each group necessarily manifests all the characteristics of earlier stages along the continuum; the four stages defined are personality disorders (psychopaths), personal illness (neurotics), integrated psychoses (manic-depressives and some schizophrenics), disintegrated psychoses (other schizophrenics). It is urged that psychological descriptions should differentiate between personality types (which are relatively unchanging and which represent personal continuity) and psychiatric syndromes (which tend to change and represent disruption and 'illness'). Between the traits of personality and the signs and symptoms of illness lie 'attitudes' which define the content as opposed to the manner of behaviour.

The implications of such an approach for clinical psychology are explored and experiments with tests which operationally define the categories (the Hysteroid–Obsessoid Questionnaire, Foulds's Punitiveness Scales and the Runwell Sign–Symptom Inventory) are reported. The results are largely consistent with the rationale.

The weakest section of the book is the final chapter on implications, which ventures into the field of aetiology (though even here Foulds's imaginative style holds interest). Laing, for example, is criticized because 'the apparent gibberish that schizophrenics talk to themselves when not aware of or not concerned about being overheard' is difficult to bring into line with Laing's notion of schizophrenia as a strategy. This kind of comment misses the point that Laing sees schizophrenia as a strategy totally involving and committing the schizophrenic, not a stratagem to be taken up or put down at will.

The classificatory focus of the book stems from Foulds's assertion that 'aetiology must wait upon nosology'. This is a chicken–egg assertion, since neither condition nor process can be well defined without reference to the other—they can be viewed as opposite poles of the same superordinate construct. Thus any exploration of process begins (if only implicitly) by defining categories of conditions; equally any description of conditions suggests a process whose parameters define the condition. It would be interesting to know how many daft classifications of blood were extant before Harvey settled on the notion of a circulatory process. Foulds's definition of personal illness as a failure to maintain interpersonal relationships urgently suggests the question 'why did they fail to maintain interpersonal relations and what inspires the establishment and maintainance of such relations?' Why should we await nosology rather than see it and aetiology as related and relatable constructs?

Foulds is too logical to take refuge in tautologies like 'constitutional predispositions' or pseudo-genetics and, for those who seek a more intelligent and better elaborated nosology, this is a very welcome book. Not least because, unlike most of us, Foulds uses his knowledge of logic rather than merely paying it lip service, and, particularly in his more discursive chapters, demonstrates that he is a writer writing a book—not merely a psychologist compiling one.

D. BANNISTER

The Junior Eysenck Personality Inventory. By SYBIL B. G. EYSENCK. London: University of London Press. 1965. Manual. Pp. 15. 3*s*. Questionnaire, 7*d*. Scoring Key. 3*s*.

The Junior E.P.I. is closely modelled on the E.P.I. (review by A. W. Heim, this *Journal*, 1965, **56**, 329–30). The published Inventory comprises 60 items, 24 each for the E (extraversion) and N (neuroticism) scales, and 12 for the L ('Lie') scale. Factorial validity of the items was established on the basis of principal components analyses for nine age-groups (7–15 years), averaging about 700 children in each group. Standardization of the chosen items was carried out for boys and girls separately, with slightly larger groups, plus a smaller, not altogether representative sample of 16-year-olds. The published norms show, for E, a slight trend up and then down through the age range for both boys and girls; for N, an upward trend for girls and a downward trend for boys, with the boys' norms considerably lower than the girls', except at age 8. For the Lie scale there is a very marked downward trend for both sexes; at age 7 the mean L scores are as high as 9 for girls and 8 for boys.

The status of the L scale seems in general to be rather dubious, since it shows substantial negative correlations (averaging over $-0{\cdot}3$) with N. The implication that stable children tend to tell more lies is perhaps surprising: on the other hand, endorsement of the relevant items may indicate not lies but actual lower incidence of the behaviour in question among stable children. The interpretation would then be that stable children are slightly better behaved (or more priggish, if preferred) than those who score high in N. (A similar point is in fact made by the author in relation to observed negative correlations, at some ages, between L and E.)

Split-half and test–retest reliabilities are quoted for all three scales, with a modal value of a little over 0·7. The position is least satisfactory for the E scale in the lower age ranges. This seems to confirm Mrs Eysenck's suggestion, *à propos* of the factorial data, that the extraversion scale is of questionable applicability before the age of 9 or 10—the evidence indeed points to a still higher age. Face validity is nevertheless high; on empirical or predictive validity there are as yet no data. The Inventory is meantime offered 'as an instrument for experimentation'. It will no doubt prove valuable for large-scale research requiring groups of children graded on the relevant dimensions. Its potential usefulness in relation to the individual case is another matter; in particular, guidance on the handling of the Lie scale would have been welcome. The author concludes: 'There is little need here here to indicate the many ways in which a valid test of personality could help the educational psychologist as well as the clinical psychologist, both in his practical work and in the conduct of experiments; there are few areas of performance or behaviour where personality is not likely to be an important variable.' This seems to imply a very narrow conception of variation in personality, one which goes little beyond a scaling of behaviour which would be very apparent to a psychologist who had in fact seen the child he was studying.

B. SEMEONOFF

Mental Retardation: A Family Study. By ELIZABETH W. REED and SHELDON C. REED. Philadelphia and London: W. B. Saunders. 1965. Pp. vi+719. 115*s*. 6*d*.

In this very large report an attempt is made at total documentation of all the relevant facts. Approximately 630 pages are given to exhaustive accounts of the families of all the probands, with pedigree charts. There is in fact enough detail in these pages to permit other workers to test out theories of their own on this material; and it is greatly to be hoped that this will in fact happen.

The field work on which these pedigrees are based seems to have been of a quality to support this great superstructure. The original investigation on which the present work was based was carried on from 1911 to 1918 at the State School and Hospital at Faribault, Minnesota, under the direction of the then superintendent. The workers concentrated on the families of patients with accessible addresses; and there is reason to think that this sample was fairly representative of the patients as a whole. The present authors then went through the records of the 549 families studied during the period 1911–18, and selected a subsample, of which all members would have an I.Q. score of 69 or lower, would all have been in the institution at some time during the period 1911–18, and would have no history of having had at any time an epileptic convulsion. The authors were finally left with 289 probands. The widest possible use for information-getting was made of social agencies of a variety of kinds. School records, including I.Q. scores, were of course of special importance and value. The first task was to follow up the probands themselves, and find out their present status and condition; the second was to trace all members of their families to the first, second and third degrees. A remarkable feature is the great co-operativeness of these relatives in answering questionnaire blanks and in helping otherwise; very few correspondents refused to answer anything. The number of individuals in these families reaching adult life, but with otherwise no useful history, seems to have been small, but does not seem to be precisely stated. For statistical purposes, all persons not positively known to be retardates are assumed to have been normal. One can therefore rely on all estimates of the frequency of mental retardation in the relatives of probands as being underestimates. The total number of persons investigated was 82,217.

The conclusions that are based upon this work are interesting, but do not go very far. One would suppose that there is a good deal more to be got out of the material available than has so far emerged. However, the following findings are of interest: Of the first-degree relatives of the 289 probands 28 per cent were mentally retarded, among the second-degree relatives 7 per cent, and in the third-degree relatives 3 per cent. For certain purposes the material was classified into 84 probands in a primarily genetic category, 55 probands in a probably genetic category, 27 probands in a primarily environmental category, 123 probands in a category of unknown causes. The incidence of retardation in the various classes of relatives differed markedly, as one might have expected, between each of these four categories. A special study of retarded persons, numbering 1450, not probands or in the immediate line of ascent or descent from probands, showed that 43 per cent of them never reproduced; this was despite the fact that very few of them (115) had ever been institutionalized. From studying the families of this sample, useful data for genetic counselling could be provided. Assortative mating between all spouses of known I.Q. was at a very high level, with a correlation of I.Q. values of +0·46. Perhaps one of the most interesting results of the investigation was the resolution of 'Cattell's paradox'. This is stated by the authors as follows: 'It has been well established that there is a negative correlation of about −0·3 between the size of a family and the average I.Q. of the children in it. The main contributions to the size and strength of the correlation come from the many small families with children having above average intelligence and a very few large families with children averaging well below an I.Q. of 100. This correlation was confirmed in our study. A possible consequence of this negative correlation would be a gradual drop in the intelligence of the population from generation to generation. The failure of such a decrease to be observed provides us with 'Cattell's paradox'. The authors were able to show that, in their sample, the phenomenon could be accounted for by the non-inclusion in the data analysed of the childless members of each generation. When allowance for this was made, it appeared that the lowest I.Q. group of persons produced the fewest children, and the highest I.Q. group the most children.

In their final chapter on the implications of the study, the authors turn their attention to the social aspects and problems. They consider that genetical causes account for most of the aetiological factors involved, with gross physical causes playing a smaller role; the part to be allotted to the purely social causes of retardation, e.g. the insufficiently stimulating upbringing given to a child by dull and underprivileged parents, can only account for a very small proportion of cases. The desirability of offering the facility of voluntary sterilization to mentally retarded persons is advocated, not least on humane grounds; for marriage provides such persons with valuable support, while reproduction adds to their burdens.

ELIOT SLATER

Factor Analysis of Data Matrices. By PAUL HORST. New York: Holt, Rinehart and Winston. 1965. Pp. xix + 730. $15.00.

The advent of electronic computers gave factor analysis renewed vigour and the number and variety of new techniques and ideas which have emerged in the past decade or so are almost overwhelming. In view of this development, the need for a comprehensive book in which the new developments were arranged in a systematic way was great. In large measure this book by Horst meets the need. Indeed, few people are better qualified than he to write it. Not only have his own contributions to factor analysis been extensive, but his very expert knowledge of matrix algebra, the ideal notation for the subject, was an essential qualification for the task.

The book is divided into six parts. The introductory part contains four chapters. In the first the objectives of factor analysis and some of its applications are set out. This is followed by two chapters on the basic principles of matrix algebra, while the fourth chapter gives a technical exposition of the foundations of factor analytic procedures. Parts II and III of the book are both concerned with 'matrix factoring methods'. The first contains four chapters dealing mainly with traditional techniques, but the author has his own individual approach to the subject which is often very original so that even the experienced reader will find much material that is unfamiliar. The chapters in Part III deal largely with computational methods. These include, in particular, variations of Jacobi-type solutions of latent root and vector problems, but there is also a discussion of other methods including the very elegant and relatively modern Householder–Wilkinson tridiagonal method. Factorial procedures are normally applied to covariance or correlation matrices, but in chapter 12 Horst shows how one may derive factor-score and factor-loading matrices direct from the raw data without ever calculating a covariance matrix.

The problem of metric is of fundamental importance in factor analytic work and in Part IV Horst discusses it in much greater detail than any previous writer has done. This part also includes chapters on 'multicategory sets' (what might be called three-dimensional type data), 'scaling' and 'image analysis'. The latter concept is due to Guttman, who claims that factor analysis should be concerned primarily with that part of each variable which can be estimated from all the other variables in the set. It is important too in that it suggests a new way of looking at the communality problem. In a modified way this idea has been used by Jöreskog in his recent and very valuable work on estimating factor loadings, though Horst does not refer to this important development.

Part V deals with the simple structure concept and with factor transformation procedures. Analytical methods of rotation, in particular the varimax method, are given special attention, but little is given to oblique solutions. It is to be regretted that no reference is made here to the more sophisticated work of Anderson and Rubin or Howe and Lawley on the testing of specific hypotheses about factor patterns when some of the loadings are postulated *a priori* to be zero. The chapters in Part VI are devoted to special topics, for example, higher-order factors, the role of factor analysis in psychological prediction, and the complicated problems which arise with multiple sets of variates—canonical analysis being a special case. Unfortunately, here again the book is a bit disappointing: basic problems concerning the conditions under which an experimenter might justifiably claim to have found the same factors in replicated studies are never squarely faced.

The book concludes with an appendix, of 119 pages, containing Fortran II programs for most of the methods of analysis presented in the text.

It is impossible in a short review to indicate the amount of ground covered in the 600 closely printed pages of straight text in this book. Many tricky side issues which have troubled students for years are clarified. For instance, it is shown how the analyst may work with raw scores, deviation scores or a combination of both. Special problems presented by variables which have true zero points, and those which have not, are discussed. Ipsative measures are dealt with. The effect of centring matrices by rows or columns is examined. The industrious student will find information on a wide variety of detailed points. But, leaving detail aside, what general assessment of this exhaustive treatise can be made?

Perhaps the main point to note is that the book is written, as Thurstone's classic books on factor analysis were written, from a mathematical rather than from a statistical viewpoint. Sampling theory is barely mentioned, and the valuable contributions of mathematical statisticians to factor analysis are virtually ignored. Each matrix of data to be analysed is taken as given, in an absolute sense. The nature or size of the sample on which the data are based do not seem to concern the author, nor do they directly enter into considerations of rank reduction. This lack of concern with sampling problems is typical of the psychometricians' approach. By ignoring the difficult problems which arise in trying to assess the stability, precision and reliability of the results obtained by his techniques, he can proceed with his model building unimpeded. But this is a dangerous philosophy and the difficulty which factor analysts in the past have experienced in trying to replicate each others' results is in large part a consequence of it. However, the user of this book will probably not go too far wrong if he confines the methods given to data obtained from very large samples.

A. E. MAXWELL

A Source Book in the History of Psychology. Edited by R. J. HERRNSTEIN and E. G. BORING. Harvard: Harvard University Press. 1965. Pp. 636. 100*s*.

It will have taken courage for a man of Boring's learning to confine himself with his collaborator in this book to more or less well-known figures. The twenty-five thousand or so English-speaking members of our unhistorically minded profession might well have made a different selection. Nevertheless, at least for a start, this was what we needed.

Perhaps because of the exciting practical applications of our science during the last generation we are apt to forget that our present has a past; and that many of the 'obvious' statements of our text-books once, and often not so long ago, involved impossible-seeming problems. As an example, one may take the now elementary fact that, while the retinal image is inverted, we see the world the 'right way up'. Even the writer's revered teacher Dodge, who died in the 30's, was puzzled by this problem that is not a problem. Any first-year student of today is liable to mutter the word 'Information' and think he knows. But which of us, not being a specialist, knew that it was the great Kepler (1571–1630) and the man whom Locke called 'the learned and worthy Mr Molyneux' (1656–1698) who posed and first solved the problem? The relevant passages are here (pp. 89–103), and fascinating reading they make. One can see what acuteness of mind and what logical care were displayed in raising and answering this question, so elementary today. One reason why psychologists should study the history of their subject is to realize what difficulties have already been overcome and how many of the great have worked to overcome them.

Herrnstein and Boring's *Source Book* is, then, a roster of great names, strewn along the march of science for more than 2000 years. Of course, by no means all were specialists in psychology, or for that matter in anything else. It is barely 200 years since the subject was finally hived off, though the Greeks were perhaps somewhat more advanced in this matter than we moderns. A look at a few of the authors is humbling. Aristotle, the inevitable for so many of the sciences. Isaac Newton, the great, on the spectrum. John Locke, who did write a book on Human Understanding, Helmholtz on vision and hearing, Kant, Hartley, Berkeley. They are represented by delicately chosen pieces, some of them translated for the first time. (I liked especially Miss Boring's work.) While, of course, we of today have the advantage of these people's labours, yet they had the advantage of the early comer. They were able to look direct at nature, neglecting in many cases what seemed the common sense of the familiar. To Cattell it seemed clear that an intelligent person is sensitive, quick, has good judgement, and so on. Therefore, in order to quantify human intelligence one should devise tests of such qualities, and add the results; and, so to speak, there you are! (pp. 432 ff.). But along came genius, unpredictable and simple. 'What we want to know is not what scientists may tell us are 'the more important processes' which characterize a person...we want a rough estimate of certain qualities in which we know (believe us!) people differ' (Binet and Henri, pp. 428 ff., 1895). Binet and Henri's method disregarded possible theoretical findings but it had the advantage that it worked.

The history of psychology contains magic names. Important passages from many of them are reprinted in this book, which is really a compendium of ideas for research; for one has the uncomfortable feeling when reading it that there are hints whose import one does not yet recognize. There are names also which one knows, but whose work one has had no chance to read. For example, Sechenov (1829–1905, pp. 308–21), whose name is held in high respect by the Russians as the precursor of Pavlov and Bechterew and who has been claimed as the father of objective psychology and thus of Behaviourism. Sechnenov has some 12 pages. There is an interesting paragraph by Weber on the 'moon illusion' (pp. 34–5). There is the great Helmholtz, master of so many sciences, for whom it is said the warning silver trumpet was blown when his carriage passed through Berlin. A personal friend of mine invented a new type of tachistoscope of which he was very proud. Years later he found the principle casually mentioned in a footnote of Helmholtz's. It was Helmholtz who said he would have sent back the human eye to his instrument maker because of its many faults. If he had been present at the creation he would no doubt have sent it back to the archangel in charge. Helmholtz is here in 25 pages or so.

And so on, with more names that ought to be dealt with but cannot for space—Titchener, Wundt, Galton, McDougall, among many. The senior psychologist should find a place on his bookshelf for Herrnstein and Boring's book. The learner who reads it may say, with the Psalmist: 'I have a goodly heritage'. May he continue with the Psalmist's prayer to maintain the lot that has fallen to him.

(The late) G. HUMPHREY

Basic Psychology. By H. H. KENDLER. London: Methuen. 1963. Pp. xii + 750. 60*s*.

Here is yet another large introductory text-book of psychology. The author's purpose is to provide 'an integrated text that not only covers the length and breadth of psychological knowledge but also provides some depth of understanding'. The book is also intended to be suitable both for students who are reading psychology only for one year and for those who intend to specialize in psychology throughout their three years. A related and laudable aim of the writer is to attract into an advanced study of psychology able students who originally intended to read psychology only during their first year.

The plan and layout of the book are based on two major assumptions, both of them questionable. The first is that, before learning any psychology at all, students should be exposed to an account of scientific method and theory in psychology: the particular brand offered is a simplified version of the old logical positivist philosophy of science, as embraced by behaviour theorists such as Spence. The second is that there is a 'core of psychological knowledge' divisible into two parts: 'basic psychological processes', such as sensation, conditioning, perception and motivation, which 'reveal how fundamental principles of behaviour operate in isolation', and 'complex psychological processes', such as learning and forgetting, verbal behaviour, problem-solving and frustration and conflict, which show how the fundamental principles 'interact'. The remaining areas of psychology, we are assured, 'can be best understood in terms of how they radiate from the inner core of psychology'. In accordance with these assumptions the contents of the book are divided into six parts: first, there are two chapters on scientific method and explanation in psychology; there follow three chapters on the history of psychology, on statistical methods and on the biological foundations of behaviour; parts III and IV comprise seven chapters on the 'basic' and 'complex' psychological processes already mentioned; next are three chapters on personality, abnormal behaviour and social psychology respectively; and the last part consists of two chapters on psychological testing and on applied psychology. The final 70 pages include a glossary of technical terms used in psychology, a list of the publications referred to in the text, a name index and a subject index; all of these are, generally speaking, accurate and comprehensive. The book itself conforms to the high standards of technical excellence that we have come to expect of American introductory text-books; the price is within the range that has become customary for such lavishly produced texts, and indeed is slightly less than that of some of its competitors.

Inevitably, complaints will be made about the omission of favoured topics. There is, for example, no reference to work on short-term memory or to the current controversy concerning one-trial or incremental learning, which is surely strange in a book which emphasizes 'basic processes' and the virtues of the experimental method so strongly; British psychologists will find odd and exasperating the solitary reference to Eysenck and the complete failure to mention Broadbent. Some readers will have doubts about the theoretical orientation, even bias, of the book, which is an interesting mixture of the ideas of N. E. Miller, Skinner and Spence, although most will admire the consistency and ingenuity with which they are used. Others may be forgiven for wondering whether a further plush and expensive introductory text-book is really necessary when so many others of similar qualities are already available. Nevertheless, here is a most impressive, comprehensive and reliable text-book, one which will certainly challenge the hegemony of such old favourites as Morgan and Munn.

G. W. PILKINGTON

An Introduction to Psychology. By J. O. WHITTAKER. Philadelphia and London: Saunders. 1965. Pp. xxiii + 631. 52*s* 6*d*.

Students' Workbook. Pp. 224. 23*s*.

This is another introductory text to add to the ever-growing library for the beginning student in psychology. It will fulfil the author's explicit intention to provide sufficient material for an introductory teaching course, and refers to much interesting current research which may stimulate good students to look further. The inevitably superficial treatment of many topics is offset by reference to specific parts of more advanced texts. The bias of this book toward social psychology is shown in the greater coverage of social psychological issues than is normally to be found in introductory texts and by the social slant to the treatments of motivation, emotion, perception and the development of behaviour.

The book is divided into seven parts. The first relates psychology to other social and behavioural studies, describes the professional opportunities for psychologists and introduces the basic physiological and cultural determinants of behaviour.

The physiological motives discussed in the second part of the book include temperature regulation, respiration, exploratory drive and drug addiction, as well as the four basic drives. A chapter on social motives provides an introduction to some of the complex variables which determine human behaviour. Part three, entitled Adjustment, contains chapters on frustration and conflict, mental health and abnormal behaviour, and readjustment of the mentally ill.

The section on learning aud retention consists of two chapters. The first, Principles of learning, deals briefly with classical conditioning and more extensively with instrumental learning. However, there is relatively little discussion of controversial issues and it might have been more satisfactory to have used some of the space given to an operant conditioning theory of drug addiction to describe fundamental problems such as the incremental character of learning. The second chapter, Human learning and retention, includes a description of programmed instruction, transfer of training and forgetting.

The fifth part af the book, The Cognitive Processes, consists of chapters on thinking (contributed by Sir Frederic Bartlett), on the sensory basis of perceiving (contributed by S. M. Luria and R. L. Sergeant) and on perception. Bartlett concentrates on three contemporary approaches to thinking. These are Bruner's experiments on concept acquisition, Posner's analysis of thinking in terms of information reduction and Jeeves and Dienes's investigations of how subjects attack the problem of discovering relations between events. This chapter contains a persisting theme based upon Bartletts' distinction between open- and closed-system thinking and the similarity between thinking and perceptual–motor skills. The second chapter in this section gives a good introductory account of the sensory basis of perception. The account is remarkable for its greater emphasis on audition than is customary. This chapter provides an excellent adjunct to the earlier biological foundations chapter. The third chapter in this section describes selected phenomena in visual perception including the development of perception (including cultural differences in perception and language) and the dependence of perception upon non-stimulus variables.

Part six, Individuality, contains a short chapter on quantitative methods in psychology (contributed by Prof. R. B. Meade) and chapters on individual and group differences and on personality. Meade's chapter introduces the basic notions of frequency distributions and their parameters, and of correlation, but its position might suggest to the student that statistical analysis was only necessary when studying individual differences. A separate section called Methods containing chapters on methodology and quantitive analysis would perhaps be a useful addendum to this text.

Muzafer and Carolyn Sherif cover social behaviour in two chapters, which should provide a sound introduction to some of the main issues in social psychological research and prepare the student to deal with more advanced texts.

While this text covers the topics needed in an introductory course the standard of the chapters is not uniform. For example, the chapters on learning and perception are more elementary than those on social behaviour, individual differences and personality and the sensory bases of perceiving. In contrast, Bartlett's chapter on thinking is more advanced than one might expect in this type of book.

The contents provide useful chapter outlines as well as chapter headings, and each chapter ends with a summary of the main points that have been covered. A glossary is provided which, though not comprehensive, should be of great assistance to the beginning student. Systematic cross-headings are used but the print size and type-face are sometimes confusing, leading to difficulty in identifying the status of a particular cross-heading.

This text is accompanied by a students' workbook containing five tests for each chapter with answers supplied at the end of the workbook. A weakness of these self-tests is that the questions sometimes allude to specific wordings from the text rather than to principles, and there is some tendency to overcue the correct answers to true/false aud multiple choice questions.

This introductory text is an interesting addition to the texts available for beginning students, differing from many in its social emphasis. This may make it particularly suitable for students who are intending to concentrate on this aspect of psychology.

D. LEGGE

Perception. By JULIAN E. HOCHBERG. 1964. Pp. x + 118. 18*s.*; *Language and Thought.* By JOHN B. CARROLL. 1964. Pp. x + 118. 12*s.* 6*d.*; *The Nature of Psychological Inquiry.* By RAY HYMAN. 1964. Pp. xi + 116. 12*s* 6*d.* Foundations of Modern Psychology Series. Edited by Richard S. Lazarus. Englewood Cliffs, New Jersey: Prentice-Hall.

It becomes increasingly difficult to know what form an introductory text-book should take. There must be few authors who have the necessary range; the very choice of chapter headings has become problematic, since several of the growing points of psychology fall between chapters; and it is even more difficult to decide how much weight should be given to the applied fields and to the necessary ancillary subjects such as physiology and statistics. If a text were to meet the needs of all introductory courses it would be of mammoth size and beyond the means of students. Prentice-Hall have very sensibly opted for different authors for the separate chapters published individually as paper-backs (the volumes are also available in hard-back), and the editor argues that it becomes possible for teachers to make up their own text-book by combining whichever volumes they find useful. This solution has many advantages, and some disadvantages, the most important being that topics which fall between volume titles tend to suffer; and the treatment of a topic is to some extent governed by the title of the book. For instance in Carroll's book, since language and thinking are treated together, there is inevitably a bias so that non-verbal aspects of thinking are underweighted.

Nevertheless, the series as a whole is of a high standard and it is already obvious that first-year undergraduates and others are profiting. Hochberg on perception has been particularly successful in producing an elementary book which steers carefully between the various controversies—analytic, holistic; nativist, environmentalist; and so on. He even introduces the reader to the 'growing points' provided by the work of Hubel and of Held. The merest introduction is given to many topics, but references are adequate and books for further reading well chosen. (One might carp that the space-saving method used for the references results in some frustration.)

Carroll's book provides a very much needed introduction to linguistics for psychology students, and, as one might expect from this author, it is lucid and surprisingly comprehensive. It is weakest on the acquisition of language, but does include an introduction to Chomsky. Although the author quotes the work of Oléron and Furth on thinking in deaf children, and, in an excellent section on individual differences in language ability, introduces the reader to the aphasias, Carroll nevertheless regards language as of paramount importance for thinking. There is however a good critical section on the few rather negative experiments which have attempted to test Whorf's hypothesis regarding the effects of differences between languages.

The third book, by Hyman, is in some ways the most interesting, and probably the most difficult. It is likely that final-year undergraduates would find it stimulating. The author has concocted a mixture of history, scientific methodology, psychological insight and deflationary criticism which is very refreshing in a book with a rather forbidding title. The last three sections on 'processing the data', 'interpreting the data' and 'communicating the results' could be read with profit by many research psychologists. Perhaps Hyman really had such an audience in mind. Some of his digs at famous psychologists suggest this, for instance when he quotes his hero, Claude Bernard, to demonstrate Skinner's lack of wisdom in using averaging techniques. Throughout the book there is an excellent, and not esoteric, choice of selections from published work to illustrate the argument.

When this series of books is compared with some recent introductory texts, the abiding impression is that it demonstrates an unusual combination of catholicism and rigour, and of historical sense and looking forward.

B. M. FOSS

A Behavior System: an Introduction to Behavior Theory concerning the Individual Organism. By C. L. HULL. New York: Wiley. 1965 (science edition). Pp. viii + 372. 17*s*.

Hull's ***Principles of Behavior*** was written nearly 25 years ago. The dust created by its publication took long to settle. On the other hand, his final essay on the application of the hypothetico-deductive method to the analysis of behaviour, *A Behavior System* (1952), made only a minor impact on psychological research and theorizing for, although it contained some important revisions to the earlier set of principles, the original array of concepts was left unchanged. Wiley have now re-issued this volume in a moderately priced edition: the book will surely find a place in the library of experimental psychologists whatever their field of interest or their theoretical persuasion.

In this book the concepts formulated in ***Principles of Behavior*** (1943) for the analysis of very simple stimulus–response situations were extended to complex forms of individual behaviour. The years 1943 to 1952 saw a systematic attack on the problems of quantifying the basic concept, ***habit strength***. Hull argued that such quantification would open the door to a radical extension of the theory to cover not only individual actions but also the variety of interactions which form the matrix of social behaviour, since it provided a common denominator, or scale, for the disparate behaviour categories which are informally gathered under the term 'response'.

The introductory chapter surveys the 17 postulates and 15 corollaries (including their quantitative expression) and their application is discussed in the subsequent ten chapters. Here Hull stated 132 theorems which he believed to be logically related to the basic postulates of the system. Amongst topics discussed are simple trial-and-error learning in a lever-pressing apparatus in terms of the differential strengthening of selected responses: the effect of massed and spaced trials on trial-and-error learning; the acquisition of simple discriminations: the learning of complex discriminations involving n-number of alternatives; the nature of generalization with special reference to stimulus intensity dimension; the problems of stimulus compounding; the analysis of higher-order behaviour like foresight, knowledge, expectancy and purpose in terms of antedating goal reactions and the stimulus consequences of such antedating goal reaction, namely ***the pure stimulus act***; the development of simple behaviour chains consisting of responses of identical structure and also of chains made up of responses widely differing in topography, e.g., lever pressing, running, chain pulling and jumping; the analysis of approach and withdrawal from objects in the environment ('behavior in space') under relatively free choice conditions; the analysis of behaviour in space under constrained conditions, e.g. maze learning; problem solving and finally the analysis of value as a special case of conditioned incentive.

As already indicated Hull did not translate the language of common sense and of pre-Hullian psychology into the language of stimulus–response learning theory. Rather he concentrated on exploring to what extent a quantified set of concepts could elucidate traditional problems. Stated differently, Hull believed that qualitative differences may turn out, upon analysis, to be differences in degree only and may thus be expressed as locations on a continuum.

A recent reviewer of Hull's contribution to the development of psychology commented that in terms of terminology (concepts) and in terms of important intensity and temporal variables, Hull kept close to Pavlov. In fact his contribution was far less original than had often been claimed. Hull's attempt to cast Pavlovian concepts into a *formal* system—postulates, corollaries, theorems—must in fact be regarded as a notable failure and has been so recognized by many of those most strongly influenced by him. To quote, 'a Pavlovian might say: let Hull drop his (drive reduction reinforcement) postulate, forget his ambitious mathematics, remember that he has a nervous system and he will be one of us' (Razran, 1965). One might add that he would also cease to be Hull and the development of contemporary psychology in the U.S.A., its theoretical orientation, its emphasis on method, quantification and sophisticated experimental design, would surely have taken a different direction. This is Hull's contribution, his greatness in the history of psychology. As an innovator he left a legacy: many have expressed gratitude and others disappointment at the size and quality of the estate, but few would gainsay the extent of his influence and the breadth of his vision.

H. M. B. HURWITZ

Attention, Arousal and the Orientation Reaction. By R. LYNN. Oxford: Pergamon. 1966. Pp. vii+118. 35*s*.

This is an eminently readable book about an important topic, relatively neglected by British workers. It effectively integrates a complex mass of data, mainly from Russian and European laboratories, and serves as a splendid introduction for those who are daunted by the tortured prose of the translation of Sokolov's *Perception and the Conditioned Reflex*, from which Dr Lynn borrows heavily. Indeed, the present volume can be regarded as a primer from which the reader can graduate to the original works on the orientation reaction of Pavlov and Sokolov and their colleagues and view them in a more comprehensive light against the background of the relevant Western activity also discussed by Dr Lynn.

Given the orientation reaction as a qualitative phenomenon, i.e. the response of the organism upon perception of a possibly significant change in the environment, the next problem is how to measure it. Most workers use peripheral physiological indices—PGR, alpha-blocking, finger vasoconstriction and head vasodilatation, but, as is usual with these methods, response specificity rears its ugly head. A PGR is not always accompanied by alpha-blocking and vice versa. Therefore single measures are not particularly valid and, although much of the work reviewed seems not to recognize this, some recent factor analytic studies are considered which appear to overcome the problem.

The phenomenon is viewed from several aspects. Ontogenetic studies show it to be subject to maturational and learning factors, depending on the individual's position on the phylogenetic scale; the further up the scale, the greater their effect. Consideration is given to the proposition that the orientation reaction is a necessary component of instrumental learning. On the evidence discussed by Dr Lynn it would not seem unreasonable to suppose that discrimination learning has not fully occurred if an orientation reaction is still given to stimuli other than the conditioned stimulus, even though there is no behavioural response. Provided one can measure the components of the orientation reaction, this suggests a criterion of learning which could prove fruitful, especially in the study of individual differences. By far the greater proportion of the book is concerned with physiological models attempting to explain the orientation reaction, its habituation and other properties. Here Dr Lynn could profitably have expanded his expositions and discussions of the various theories. One is presented with a catalogue instead of a review, which at times is frustrating. The catalogue is stimulating though, in that it makes one go back to the original papers—which could be what Dr Lynn intended. The work considered indicates that there are at least two stages of the central nervous system involved in the production of an orientation reaction—a subcortical arousing mechanism and a cortical evaluating system; the hippocampus is probably important also, together with the afferent pathways; there are feedback loops everywhere and the situation, as might have been guessed, is far from clear. Dr Lynn gives one hope, however, that order is almost ready to spring from chaos.

The most fascinating chapter is the last, where individual differences are considered, albeit fleetingly due to the dearth of information on the subject. Here is a positive goldmine of research projects. One is left with the urge to rush out and solve the problems left hanging so tantalizingly in this book. Investigations concerning personality correlates of the orientation reaction, and physiological measurements taking into account the general activation level of the organism—a factor which appears to have been overlooked—are two of the most important which spring to mind. In fact, one of the good things about this book is the number of experiments it suggests; it should be on the shelf of all persons interested in human behaviour, especially those concerned with personality and learning.

It may seem churlish to criticize at this stage, but the assertion on p. 77, that the *cause* of distractibility in mentally defective children is that orientation reactions are given to intense irrelevant stimuli, is a bit hard to swallow. The orientation reactions are a *manifestation* of distractibility, but hardly a cause. However, this is a relatively minor point; on the whole Dr Lynn keeps up the high standard set by Gray's volume on *Pavlov's Typology*, in this series of Monographs in Experimental Psychology. He has made a not insignificant contribution to the effort of linking East and West in the scientific investigation of behaviour.

W. I. HUME

Brain Research. Vol. I, No. I, January 1966. Elsevier. Issues to appear monthly in two volumes per year. Subscription £5. 12*s.* 6*d.* per volume.

The stated scope and purpose of this new journal is to provide a medium for prompt publication of articles in neuro-anatomy, neurochemistry, neurophysiology, neuroendocrinology, neuropharmacology, neurocommunications, behavioural sciences, molecular neurology and biocybernetics. It is intended that it will contain not only research reports but also critical reviews. The journal forms a companion publication to the series of books *Progress in Brain Research* and it is to be supposed that psychologists may gauge their possible interest in the new journal from the value they have found in the 'Progress' series of books.

The first number of the journal contains no papers of a directly psychological nature, although several will be of use to those working in the field of physiological psychology.

It is to be hoped that the presence of some well-known behavioural scientists on the editorial board will ensure the publication of papers which will be of more than peripheral interest to psychologists.

PETER VENABLES

Adaptation-Level Theory: An Experimental and Systematic Approach to Behaviour. By HARRY HELSON. New York: Harper and Row. 1964. Pp. 732. 81*s.*

The notion of adaptation level had its start in 1924 when the author of this ambitious book, Prof. Harry Helson, together with D. B. Judd, found that in some chromatic conditions 'the eye may supply complementary colours when there were no specific stimuli for them'. This has its origin in Goethe's coloured shadows, and its striking fulfilment in E. H. Land's two-colour additive projection demonstrations. These appear to violate the classical three-colour experiment of Thomas Young, which showed that for simple lights three colours is the minimum to give by mixture any spectral hue and white. Helson was aware 40 years ago of the basic facts so dramatically shown in the recent Land demonstrations, and he explained the 'subjective' colours in terms of his now celebrated adaptation-level concept. This is a quite complicated notion, for it implies a double effect: reduced response to 'dominant' stimulation and heightened response to 'complementary' stimulation. Between these there is a 'neutral region', this being an equilibrium condition which is variable and is supposed to represent the mean of all stimuli, past and present, affecting behaviour.

If indeed sensations and sensory judgements are due as much to the current setting of the neutral region as to the prevailing stimuli then this notion of adaptation level is of first importance, for a mere statement of stimulus conditions could not be sufficient to describe even the most basic 'physiological' sensory facts. Psychologists do tend to think this way, citing such well-known effects as the size–weight illusion, but if Helson's claims are justifiable these added 'psychological' effects might be precisely quantified and built into the very foundations of sensation and perception, rather than relegated to the low status of annoying complications.

Years ago adaptation-level theory must have looked bright with promise. Not quite stimulus-bound, yet quantifiable; genuine phenomena to shake the simple faith of sensory physiologists that the important secrets lie in the periphery, so that the too-difficult brain can safely be ignored. Adaptation level could indeed have been a banner to march under. But the trumpets were too few and strangely muted, as though uncertain of their foe or where to pitch the tents for battle. Adaptation level became accepted rather as a complication in interpreting experimental data than a basic fact of sensory life, to be used for integrating and interpreting the major phenomena of perception.

In this book Helson does not, as one might have expected, so much consolidate his initial position with careful quantification in simple experimental situations but rather tries to extend the notion to cover, in a general way, the whole of psychology. The section on perception, which one would have thought crucial, is comparatively short and disappointing in its limited discussion of issues clearly important to his cause, while he deploys and dissipates his forces on a host of issues in learning, thinking, group behaviour and individual personality characteristics. If the first battles had been decisive victories this could have been a *tour de force*, but doubts remain at the most fundamental level of the power of the concept to deal effectively with the simplest issues, let alone the whole field of human behaviour. Granted that it is a factor to be considered, yet it never gives the insights of a truly powerful unifying concept. We are uncertain whether it is even quite sufficient for the Land demonstrations—which should reveal it in purest form—so why should we expect victories in learning, personality and the rest? Important and exciting problems are reduced to dull reiteration. The idea, so potentially important, seems to intrude when made a major element in complex situations in which no doubt it has a place—but not, surely, as major-general. But the book will be welcomed as a statement of a brave attempt to achieve some kind of unity, by developing an idea and applying it to difficult problems.

R. L. GREGORY

Eye and Brain: the Psychology of Seeing. By R. L. GREGORY. London: Weidenfeld and Nicholson. 1966. Pp. 251. 14*s*.

There would seem to be three important characteristics of a popular introductory text on any psychological topic: first, that it should attract the attention of the general reader; secondly, that it should be readable and comprehensible; and thirdly, that it should present a fair coverage of the main features of that topic.

Mr Gregory's book, *Eye and Brain*, admirably demonstrates the first two characteristics. Because a very large initial issue has been printed, the publishers are able to produce a most attractive book at a very moderate price. It contains some beautiful reproductions of coloured pictures, many photographs and excellent diagrams. Again, Mr Gregory has written a most readable book, in a clear and stimulating manner, which includes a wide background of interesting general information.

But as regards the third of the above characteristics, the book is less satisfactory. Clearly it must be selective rather than exhaustive, and there is no reason why it should not concentrate upon particular aspects of vision. But these should be significant rather than trivial. Mr Gregory does indeed begin with a lucid and instructive outline of the physiology of vision; and he also provides an excellent discussion of space and distance perception. But he seems to omit all consideration of the perception of forms and objects, though without this there could be no space perception. Indeed, perception of form or configuration is undoubtedly fundamental to all perception. In vision, seeing edges or even outlines is only the first stage; invariably, in ordinary perceptual situations, they cohere to produce form. The findings of Hubel and Wiesel, significant though they may be, do not explain this. Perhaps the work of the Gestalt psychologists may seem *vieux jeu* nowadays; but their fundamental tenets still hold good. Moreover, there is much new and interesting work on the characteristics of form and object perception, for instance on redundancy and its differing roles in discrimination and identification.

Furthermore, it would appear that form perception plays an integral part in perception of visual illusions, to which Mr Gregory devotes so much attention. Yet he seems to dismiss form perception rather cavalierly under the term 'confusion theory', in favour of the 'constancy scaling theory', the evidence for which is by no means universally positive. Another important topic, children's perceptions of illusions, which Piaget and Vurpillot have shown to be highly relevant to perceptual learning and development, is not mentioned.

In recent years it has become apparent that visual perception is a function not only of retinal and cortical mechanisms, but also of the subcortical reticular formation. The interaction of these mechanisms and the part they play in determining what is perceived from moment to moment is by no means clear. But there can be little doubt that 'arousal' is involved in attention, vigilance, sensory (or rather, perceptual) deprivation, and even perceptual defence. Mr Gregory might have contributed a useful up-to-date discussion of this important new work. Unfortunately, he has scarcely touched on it, except in connexion with space travel and arrival on the moon. A consideration of some mundane applications would have been appropriate.

There can be no doubt that the general reader and the student beginning a course on psychology will derive interest and profit from reading this book; but a differently orientated selection of topics might have taught them more.

M. D. VERNON

Sensory Restriction: Effects on Behavior. By DUANE P. SCHULTZ. New York: Academic Press. 1965. Pp. 216. $7.50.

To anyone whose knowledge of sensory deprivation studies is limited to the dramatic McGill experiments carried out in the early 1950's, it may come as a surprise to find almost 250 references cited in this present survey. The reason for the popularity of the topic is not hard to find. Where experiments can throw light on brain mechanisms on the one hand, yet yield important practical information on such activities as motorway driving, task monitoring and manning of Arctic stations, submarines and space-craft on the other, there should be no shortage of willing sponsors or researchers. Apart from its sheer volume, the other striking feature of the literature is the great diversity of results reported. This divergence from the relatively clear-cut conclusions of earlier studies is due partly to the increased rigour of current methods and partly to the wider range of situations and subjects employed.

The author of the present volume organizes his survey of the current position by presenting his own theoretical framework and then systematically examining the literature to see how far the data fit predictions from his model. His basic argument is that the cortex requires a certain level of activation in order to function optimally. This activation is provided by variation in exteroceptive stimulation and is monitored by the reticular formation. Balance in the amount of sensory variation relayed to the cortex is akin to homeostasis and is termed 'sensoristasis' by the author. Conditions of gross sensory restriction or overload will upset the sensoristatic equilibrium and it is the effects of the former condition which are considered here.

This approach provides a plausible, though hardly novel, theoretical framework and the author proceeds to examine the effects of sensory restriction on physiological, cognitive, perceptual, motor and affective variables. He also includes a discussion on the very large individual differences reported (one of the few points of complete agreement in the area) and a chapter on the effects of social isolation. The results of more recent studies show the effects of sensory restriction to be much less severe than was formerly supposed—the situation was even found to be beneficial for some learning tasks and in some cases of psychiatric treatment. Other striking points which emerge are the importance of social factors and the care which must be exercised in generalizing from laboratory to real-life conditions. There is some evidence, for example, that knowledge of length of time to be spent in isolation and also the freedom to terminate the experiment are important factors which may be missing from real-life situations. It is also interesting to note that studies on groups of people isolated in simulated fall-out shelter conditions have shown few effects due to social isolation, though we may perhaps question the author's use of the term 'optimistic' to describe this result! In general, the experimental conclusions considered are in agreement with the author's main theoretical predictions.

The book seems comprehensive in its coverage and is certainly written in an agreeable and lucid style. A helpful feature is the table provided at the end of some chapters comparing the main conditions and results of experiments in a particular area. Of the three theoretical approaches generally adopted to studies in sensory restriction—psychoanalytic, psychological, and neurophysiological—the author favours the latter, but, as the facts and arguments are presented in an unbiased fashion, the survey should also be useful to those who are more sympathetic towards other approaches. To anyone who wants a fairly short, readable account of work in this area, this book may be recommended with the usual reservations which apply to surveys of a rapdily expanding research area. The author himself points out that, 'At the rapid rate at which the literature is increasing, this book will enjoy a relatively short, but hopefully useful, life.' In any event it is a most useful gathering together of fact and theory in an important branch of experimental psychology.

F. I. M. CRAIK

The Spiral After-Effect. By H. C. HOLLAND. Oxford: Pergamon. 1965. Pp. xi+128. 35*s*.

Long ago, in the first quarter of the nineteenth century, Purkinje made the first scientific report on the visual after-effect; a few years later, Plateau devised the rotating spiral now wrongly ascribed to Archimedes. Yet more than a century was to elapse before the spiral became widely used and a conventional part of the psychological laboratory. This mid-twentieth century boom was due principally to the perpetually misguided enthusiasm of the clinicians, who, decked in the scientific panoply of stop-watch and retired gramphone motor, saw in the spiral 'test' that chimerical lure of all test constructors—the 'valid, reliable and two-minute test of brain damage'.

By the beginning of the present decade, the contradictions in results of experimental studies, only partly explained by the diagnostic gallimaufry of the standardizing populations and the naïveté, faith and blind hope of some of the investigators, convinced all but the scientifically myopic that as a test the spiral after-effect had outlived its usefulness. Useful it had been, not, alas, in differential diagnosis, but in drawing to the attention of the experimental psychologists what Holland calls the unbearable ambiguity of its many variables.

The second half of the book is taken up in defining and explaining these variables, which fall into three groups referring to the characteristics of the stimulus, method, and subject respectively. It is encouraging to note that, if the former two groups of variables can be held constant, the spiral after-effect has potential value in determining individual differences. It would still be rash at this stage to deduce that such differences referred to brain damage, extraversion, intelligence, or any of the other factors which seem to influence the phenomena, but there are reasonable grounds for believing that immense potentiality still exists for exploring these variables and using them in the applied fields of psychology. The drooping spirits of clinicians will be revived by the author's discussion on the use of the spiral after-effect to explore the localization of brain function and dysfunction, while his critique of the SAE–extraversion relationship adds another buttress to the much-assailed Eysenckian tower of cortical excitation.

An extensive historical review, which occupies the first half of the book, provides with tabulated data a concise statement of the more important findings to date, and is the first complete review of experimental work on the spiral after-effect to be published. The experiments of Wohlgemuth, conducted sixty years ago, are summarized in some detail and commended for their accuracy and inventiveness: it may be salutary to reflect that much of the similar work undertaken in our own times compares unfavourably with that of the late nineteenth-century experimentalists.

Dr Holland, as the leading authority on the subject, is especially well qualified to write this book, which makes a valuable and unique contribution to psychological literature. Himself a meticulous worker in experimental psychology, he has added to historical literature as both historian and as one of its principal figures in this field, at the same time providing an essential handbook which adds to the status of SAE measurement and brings a useful tool into many diverse areas of applied psychology.

L. R. C. HAWARD

Readings in the Study of Visually Perceived Movement. Edited by IRWIN M. SPIGEL. New York: Harper and Row. 1965. Pp. ix+347. 56*s*.

From the biological point of view movement perception is one of the most fundamental aspects of vision yet it often receives scant attention in text-books of experimental psychology. To help remedy this deficiency and 'to provide some indication of the direction and scope of research into the visual perception of movement' Dr Spigel decided to publish the present volume of readings. The book is intended for advanced undergraduate and graduate students. To increase its usefulness as supplementary material for courses in sensation and perception or experimental psychology, papers have been selected, wherever possible, from 'less readily accessible sources'. Everything hinges on what is meant by 'readily accessible': so far as psychology students in this country are concerned it seems unlikely that they would have any difficulty in obtaining 13 of the 18 papers that have been reprinted in this volume. This detracts somewhat from the book's value but it might nevertheless be convenient and useful to have 18 papers on movement perception selected for you by an expert in the field and brought together in a single volume.

Dr Spigel's selection is divided into three parts (Real Movement, Apparent Movement and Physiological Bases) and covers a period of about 30 years, starting with two well-known articles by J. F. Brown on movement thresholds and the perception of velocity that first appeared in 1931. Most of the psychophysical papers merit serious attention (there are too many authors to mention names) although some of them make rather dull reading. The most stimulating part of the book is the physiological section, which contains two microelectrode studies of the higher visual centres by Hubel and Wiesel and by Ursula Grüsser-Cornehls, O-J. Grüsser and T. H. Bullock. This work is of exceptional interest in showing that there are specific detectors for movement in the nervous system. In view of its outstanding importance it is a pity that the Hubel and Wiesel paper appears in an abridged form with the sections on methodology, binocular interaction, ocular dominance and cortical cytoarchitecture left out and with incorrect figure references in its tables. If it was necessary to economize on space it would surely have been better to have reprinted only one of the two lengthy and well-known articles by J. F. Brown that together occupy as much space as the whole of the physiological section of the book, rather than mutilate the article by Hubel and Wiesel.

By leaving out one of Brown's papers and the 30-page Introduction, which adds little to the value of the book, it would have been possible to extend the physiological section to include a sample of recent important work on retinal mechanisms. Besides more physiology one would have liked to see more comparative behavioural studies to provide the reader with a broader biological setting for studying movement perception. Despite its shortcomings this book will give students a better idea of the scope of research in this field than they can get from their textbooks. Dr Spigel has succeeded, therefore, in one of his main aims.

G. W. GRANGER

Behaviour in Uncertainty and its Social Implications. By JOHN COHEN. London: Allen and Unwin. 1964. Pp. 207. 30*s*.

This is Prof. Cohen's third book on decision making and uncertainty, and this time he has wandered even more widely than before through history, literature, and the current social scene in an attempt to illuminate his theme. Sometimes, however, one feels that he has strayed into topics that, though fascinating in themselves, seem to be only remotely relevant. His last chapter, for example, gives a succinct survey of the different methods that have been used for divination: apparently he feels the subject to be so important that it is supplemented with an appendix listing about sixty different techniques, running from alectromancy (using a cock) to tyromancy (using cheese), while another appendix gives a long extract from De Quincey's *Confessions* on the topic. But apart from commenting that 'divination may be regarded as an archaic method of decision making', the chapter seems to be more suitable as a digest of the subject of divination for a popular encyclopedia, than as a contribution to the study of the psychology of decision making.

The same criticism might be made of another chapter devoted to suicide. The statistics and the anecdotes are again fascinating, but it seems to be rather stretching the general theme of the book if the suicide is considered as 'engaging in a wager, for he is hazarding his life in the hope of proving that he is cherished by the gods'. This thesis, which is presumably the justification for including the chapter in the book, is hardly substantiated by the wealth of facts summarized by the author. Apart from indicating the bewildering number of factors which must be taken into account when studying suicide, it is never shown how considering even some kinds of suicide as a form of wager gives one any particular insight, or how one might be able to explain cases which are inexplicable on the basis of other theories.

More pertinent is a lucid and incisive chapter on the risks and hazards of the road, appropriately entitled 'Gambling with life on the road'. In particular the by now well-known facts concerning the effects of alcohol on driving are summarized and placed within the wider context of social planning and administration, while the original point is made that 'alcohol brings disasters on the road less because of lack of skill than because of defective judgement in relation to skill'.

A chapter on risk taking in sport reports an amusing study of the accuracy with which footballers can estimate their chances of scoring when placed at different distances from the goal. There was little difference between West Bromwich Albion and Manchester United, while both showed the well-known phenomena of overestimating small and underestimating large objective probabilities. On the whole, however, they were surprisingly accurate, and naturally better than either Manchester University or Grammar School teams.

The report of the author's experimental studies of gambling is perhaps the most interesting section of the book. In one investigation people had to choose which of several different methods of drawing from a lottery they would use. The work is important, since the results clearly show that the manner in which statistically equivalent randomizations are produced strongly affects people's estimates of their chance of success. A consequence is that most of the mathematical models of decision making under uncertainty are inadequate, since they do not take into account the particular method which is used to produce chance outcomes. With these models the experimental procedure is only specified in terms of objective probabilities, the perceptual aspects of how these probabilities are produced being ignored. Prof. Cohen's results clearly show that two possible ways of winning a prize in a lottery, though having objectively equal chances of success, do not appear to be equivalent. His studies of the phenomenology of chance are a most original contribution to the subject, but unfortunately the author's tendency to display encyclopedic knowledge, almost for its own sake, reasserts itself in the next chapter on the history of gambling and lotteries.

In general the book is packed with information for the curious, but the reader for the most part will have to work out for himself the details of its psychological and social significance.

A. R. JONKHEERE

The Organization of Research Establishments. Edited by Sir JOHN COCKCROFT. London: Cambridge University Press. 1965. Pp. 275. 63*s*.

In his introduction Sir John Cockcroft twice indicates that the purpose of the book is 'to discuss the factors which make for creativity and productivity in a wide variety of establishments'; this objective does not seem to have been communicated to the contributors and indeed Sir Charles Harrington in his chapter on the need for medical research describes the book more accurately when he says (page 91) 'the stated purpose of this volume being to describe the organization of research institutes'. For this reason the book is a disappointment, but it nevertheless contains a number of interesting discussions of the history, organization and achievements of research establishments. The establishments concerned include the National Physical Laboratory, the Royal Aircraft Establishment, the Atomic Energy Research Establishment, the National Institute for Medical Research, the Medical Research Council's Social Psychiatry Research Unit, one of the Agricultural Research Council's Institutes, the Empire Cotton Corporation, the Glaxo Research Organization, the British Iron and Steel Research Association, British Railways Research and Development, the Bell Telephone Laboratories, the Rutherford High Energy Physics Laboratory and CERN.

Some interesting observations are made which include Lighthill's comment on page 53 that there are almost no limits to the benefits that size can confer, given homogeneity of aim and organization designed to avoid fragmentation. Unfortunately this point is not taken up by the other contributors and indeed there is an implied contradiction in the descriptions that some of the other contributors give of the efforts that have to be made to avoid scientists becoming too isolated intellectually. This point is taken up by Sir Charles Harrington, but unfortunately he can only conclude that the facilities he makes available give 'no reason why any member of staff should remain in ignorance of what is going on outside his own division'.

Sir Aubrey Lewis, in his description of the Social Psychiatry Research Unit, shows concern with the problems of keeping research workers' interest fresh, but offers only absence of organization as a solution, and as Sir Edward Bullard points out in his summary 'accounts of the smaller organizations concentrate mostly on the work done and say little about the organizational problems of getting people to do it...In the large organizations the organizational problems loom much larger and become one of the main concerns of the management.' One or two contributors, Sir Harry Jephcote in particular, are concerned with the problems of how to terminate a line of research, which is indeed one of the most ticklish problems which faces a research director; and Jephcote makes a very wise observation to the effect that it may be better to let people go on until they themselves see that the project will fail, before winding up. The industrial contributors show an awareness of the need to evaluate the research in financial terms, a consideration which is very much less prominent in the account of the research organizations in the public sector. It is left, however, to Dr Inglis of British Railways to offer us the bland assurance that 'the objective review which is made of all research programmes on the basis of these primary aims *ensures* the efficient and productive use of the research facilities of British Railways' (reviewer's italics).

The best contribution to this book is that of Dr Fisk of the Bell Telephone Laboratories, who makes some very telling points on the requirements for academic work in industrial research establishments, on the relationship of the cost of discovery to the cost of exploitation, of the need for a comprehensive training programme for graduate entrants; and people in universities would be delighted with his comment 'it cannot be wise to invest happily in people and be parsimonious about the physical things or technical or clerical assistance they need'. Dr Fisk is the only contributor to point out the importance of a performance appraisal system for research workers. This indicates what is really missing in this book. There really is no description of what makes for individual creativity and performance, and, although Sir Edward Bullard may be right to say that 'the first consideration in the organization of a research establishment is to foster the right climate of opinion—perhaps this is more important than the quality of the staff', this is not a fruitful generalization for detailed analysis. For this, one would have to turn to a recent publication *Administering Research and Development* by Orth, Bailey and Woleck, (Tavistock Press, 50*s*.), a much better buy in pence per pound desk weight.

ALBERT CHERNS

Widening Horizons in Creativity. Edited by CALVIN W. TAYLOR. New York: Wiley. 1964. Pp. xix + 466. 68*s*.

The articles and discussions contained in this volume are the fruits of a conference held at Utah in 1962, the fifth of a series, on the subject of 'creativity'. They show that a large amount of empirical research has been carried out on this topic during the last decade—also that we still await a breakthrough, a momentous discovery which will revolutionize the study of original thought. The problems of prediction or explanation in this area are, it seems, roughly as intractable as they ever were. And, indeed, now that the first flush of their enthusiasm is waning, workers in this field are beginning to change their research strategies. One notes especially the insistence of Barron and Leary on the experimental modification of individuals' behaviour—using, particularly, hallucinatory drugs. Their aim is not to differentiate the 'creative' from the 'non-creative' but to induce in a wide range of individuals more 'creative' states of mind. The question of whether these states of mind are closely related to the practical business of innovating in the arts or the sciences is left open. The states of mind are explored for their own sake.

Barron and Leary's contributions are not without their philosophical or even mystical overtones, and, superficially at least, seem poles apart from the more painstaking predictive studies of 'creativeness' in a variety of situations: school, college, laboratory. However, nearly all the contributions to this conference seem in some form or other to embody a concern for the modification, the improvement rather than the measurement, of human performance. A further point, brought out by Hyman, as well as by Barron and Leary, is that a person's intellectual performance may depend not solely on his abilities, but on those of his abilities which he feels free to deploy. And it seems that this freedom, in its turn, is dependent on quite subtle aspects of an individual's 'self-image'. Thus, a rigid, cautious man who normally produces few word associations may produce many more if he is asked to 'pretend' for the moment that he is someone else—a person who is fluent, relaxed, easy-going. Harmon remarks this general change of emphasis in words which might well apply to the book as a whole: 'In summing up and characterizing these rather heterogeneous scraps of evidence, the first point concerns human learning potential. There is excellent evidence of far more learning ability—mathematical, verbal, aesthetic (and, I would add, social–emotional)—than our psychometric philosophy has characteristically assumed.'

L. HUDSON

Studies in Psychology. Scripta Hierosolymitana vol. XIV. Edited by R. R. EIFERMANN. Jerusalem: Magnes Press. 1965. Pp. 43*s*.

This volume is a collection of theoretical and empirical work done by members of the Department of Psychology of the Hebrew University in Jerusalem. The volume is divided into four sections: studies on development; social behaviour and personality; tests and measurements; cognitive processes. Each of these sections includes papers widely differing in approach, quality and importance. Ortar's paper describes a follow-up study concerning the validity of entrance examinations to the Psychology Department which seems to have strayed by accident out of the departmental files. Guttman's paper, on the other hand, is outstanding and deals with a faceted definition of 'Intelligence'. The main point of his paper is to convince the reader that 'it is feasible to make certain definite statements about what an intelligence test is testing, provided a clear facet design is used for item construction'.

The same divergence in quality appears in the first section on development. Here J. L. Gewirtz's study on smiling and the study by S. Kugelmass and his associates on moral judgement are on a much more sophisticated level of thinking and methodology than the one by Kohen-Raz on conceptual thought at early school age. Gewirtz's study compares the course of the development (1–18 months) of smiling in different environments (e.g. institution, kibutz). The paper also provides a theoretical analysis of the development of this early social response in 'learning theory' terminology. The paper by Kugelmass *et al.* investigates the development (11–17 years) of 'intentionality' as a criterion in moral judgement. It is presented as a partial test of some of Piaget's notions. Unlike many other studies in this field this study is well designed and controlled. One of the more interesting findings of the study is that 'intentionality' develops continuously with age.

The section on social behaviour and personality includes three studies:

(*a*) Herman and his associates have tried to investigate the relationship between 'involvement' with Judaism and attitudes to the Eichman Trial. Unfortunately the study suffers from poor methodology and the very loose presentation does not take full advantage of what the Eichman trial had to offer regarding public opinion changes.

(*b*) Minkowitz's study is of the psychoanalytic type following the tradition set by G. S. Blum. The results however challenge the Freudian assumption of the superego as a unitary function.

(*c*) The study by Raven and Leff investigates the effects of 'partner's strategy' as well as 'ideological upbringing' on the choice of strategy in a two-person (non-zero-sum) game situation. Contrary to expectation kibutz youth tend to follow an essentially competitive strategy rather than a cooperative one.

In Section 4 there is a paper by R. Eifermann which is related to the Bruner *et al. Study of Thinking,* and another by I. Schlesinger which demonstrates that when redundancy is introduced in the absence of stimulus 'noise' it affects a card-sorting task adversely.

On the whole the volume makes worth-while reading. Its especial interest, however, is the exploration of some special features of the Israeli community (e.g. kibutz life) which are of great interest to psychologists and sociologists in general.

M. ROSENTHAL

An Introduction to Mathematical Learning Theory. By R. C. ATKINSON, G. H. BOWERS and E. J. CROTHERS. New York: Wiley. 1965. Pp. xiii+429. 75*s*.

This addition to the rapidly expanding range of books on mathematical psychology is aimed principally at students and research workers who possess only limited mathematical training. For once, here is a book which succeeds in presenting its mathematics simply, while taking mathematical derivations into considerable detail. The realm of the book is the application of mathematics to the psychology of learning but it is by no means restrictive in its approach. The chapters in general deal with separate psychological topics—for each of which one or two models are formulated and studied in detail.

The first is concept identification, which is treated from the hypothesis-testing approach. Next the one-element model for paired-associate learning is compared with its antithesis—a simple linear model. These are good models for comparison since both have only one parameter, yet one is incremental and the other all or none. Their combination in the random-trial increments model is also discussed. Choice behaviour is examined with a clear account of Luce's choice axiom. However, despite the obvious application to learning, the illustrations given are those from ranking and preference experiments. There follows a random walk model for vicarious trial-and-error which is subsequently extended to cover choice times. In this chapter the authors vacillate between a desire to review various models and their intention to demonstrate the manner in which predictions about choice times can follow from particular representations of the response process. In including a chapter on signal detection and probability learning the authors present a model which will account for either situation. The signal detection model is orientated towards learning theory—subjects modifying their bias after incorrect reports according to a linear process. With weak signals the signal detection situation becomes more one of probability learning and the two cases are combined in the last section. The treatment here is of very limited material but intentionally so, for no review of theories is attempted while the derivations are carried into some detail—the emphasis as elsewhere in the book being as much on developing techniques as on theoretical discussion. A return to strict learning theory occurs in chapter 6 when avoidance conditioning is investigated. The approach here is to seek a quantitative model for the data which may be incorporated into Mowrer's two-process theory of avoidance conditioning. There is an interesting example of the use of the Monte Carlo simulation method to compare randomly produced with empirical data for Bush and Mosteller's two-operator linear model.

Mathematical learning theory predictions are derived for individual behaviour in social situations. It might be doubted that learning theory is applicable here, but, if learning is defined simply as the effect of reinforcement on response probability, then in social situations the reinforcement depends on the subject's behaviour and that of others. The application of learning models to social behaviour is demonstrated with a two-person collective versus self-interest game. Both this and a model presented for economic behaviour in an oligopoly involve Markov chain processes. Chapter 8 treats a specific theory rather than a problem—Estes's stimulus-sampling theory. Mathematical predictions are made for the N-element pattern model and are applied to a probability learning experiment.

The approach in most chapters is to define the problem and formulate a model, from which theoretical expressions are derived for the parameters. These are then estimated and tested against the data. The models selected are suitable for simple exposition while illustrating varied mathematical techniques. It must be emphasized that this book is not a review of mathematical models of learning, nor of empirical evidence for the models chosen. Rather does it demonstrate how models may be used in learning theory—the models chosen reflecting the personal interests of the authors. However, for comprehensiveness there is an adequate bibliography. The usefulness of the book lies largely in the way in which it develops mathematical techniques so that these may be applied elsewhere by the reader. Markov chain models abound, but the total range of techniques introduced is extensive. Despite the inclusion of much that is purely mathematical the text is highly readable.

ADRIAN J. SIMPSON

Human Factors Evaluation in System Development. By G. MEISTER and D. RABIDEAU. London: Wiley. 1965. Pp. xii + 307. 75*s*.

This book contains a couple of parts. The first is devoted to man (the chapters are headed 'Analysis and Evaluation' and 'Human Evaluation of System Products'); the other part is concerned with data analysis and system performance. The authors use an imaginary system development project (rather closely related to real projects in aerospace development programmes) as an illustrative theme. There are numerous tables showing what is superior to what, for example in what respects a man is superior to a machine and vice versa; there are charts showing how the 'human factors' investigation is conducted and how the functions of a system can be analysed. But all of these heuristic aids fail to produce a very coherent image of the field. The authors say that 'Human Factors' is a technology rather than a science, although it may become a science. They are probably right. But, as presented, it has insufficient content even for a technology.

Part of the trouble is that the authors' perfectly defensible suspicion about the practical applicability of results gleaned in the laboratory leads them to assume an undue ignorance of human characteristics. This point of view is neither consonant with the authors' (avowedly pragmatic) attitude nor in keeping with the facts. Something *is* known about the human being as a system component; for example, quite a lot is known about the data-processing capabilities of man from the work of L. Augenstein, J. G. Miller, and Swets and Tanner in America and the A.P.U. in Britain; something *is* known about groups of men, for example from the work of Bavelas, Lee Christie, and the Romes' elaborate man computor system, Leviathan. None of this, or any comparable work, is mentioned.

The system theoretic material is scanty though much more illuminating. The summaries of analytic methods and forms of measurement are very useful but some important issues, such as the design of redundant networks, are not considered. Feedback control in systems, with or without man as a component, is touched upon but its theory is not developed.

Because of these blemishes the book is neither a psychologically nor cybernetically authoritative account. But it may prove valuable to the psychologist as a fairly detailed description of what goes on from day to day in the 'Human Factors' department of a large system development programme.

GORDON PASK

Supplement to Tables for Testing Significance in a 2×2 Contingency Table. By B. M. BENNETT and C. HORST. London: Cambridge University Press. 1966. Pp. iv+28. *5s.*

The original table compiled by D. J. Finney and published in *Biometrika* has been much used by psychologists, as has the first extension compiled by R. Latscha. These two tables were published by the Cambridge University Press for the Biometrika Trustees in 1963, together with an extended Table of 5% and 1% significance points up to $A = 40$, $B \leqslant A$, which had been compiled by Prof. Bennett and P. Hsu. Based on Mr Hsu's IBM program, the present *Supplement* extends the Table of 5% and 1% single-tail significance points for $41 \leqslant A \leqslant 50$, $B \leqslant A$. Finney and Latscha made two-tail tests of significance possible by also giving 2·5% and 0·5% points. In welcoming the new extension, may one hope that the next step may be to persuade the computer to yield the other two sets of points up to $A = 50$?

ARTHUR SUMMERFIELD

Porteus Maze Tests: Fifty Years Application. By STANLEY D. PORTEUS. Palo Alto: Pacific Books. 1965. Pp. 320. $7.50.

In this book Porteus traces the development of the Maze Test from its first uses with feeble-minded children in 1915 up to the production of a second parallel form of the test in 1959. Porteus deals in turn with the results of studies in which the Maze Test has been used in assessing cognitive changes following psychosurgery or treatment with ataractic drugs, differences in performance between social offenders and normal persons and between the sexes, and also the test's usefulness in the industrial setting. The latter part of the book deals with the Maze Test's contribution to the study of cognitive differences between ethnic groups. In his final chapter Porteus gives the latest instructions and scoring systems for the test, and the appendices contain all three forms of the test plus tables of test quotients and examples of scoring.

Porteus considers his test to be one not only of intelligence but also one of foresight and planning. The main criticism of this book is that he fails to delineate in any clear fashion just what the test really is supposed to be measuring. There appears to be no extensive study reported in the literature which answers this question. A wealth of studies is cited in the book purporting to show what the test measures, but they become a little confusing after a while.

The book itself is written in a rather peculiar style. It is a mixture of the historical novel, test manual and social anthropological report, peppered with the most extraordinary passages of preaching and scolding. Many readers may find this quite irritating and distracting.

D. C. KENDRICK

The Gibson Spiral Maze Test. London: University of London Press. 1965. Pp. 12. Manual 4*s*. Test cards 4*d*.

Maze tests have not always received the attention they merit. This is seen from Gibson's work with a simple spiral maze which takes about 2 minutes to administer.

He uses a pencil and paper test, boldly printed on a 10 in. × 12 in. card. A person is asked to start at the centre and to get out as quickly as possible without touching the side lines or any of fifty-six obstacles placed in his path. There are no wrong turnings. The time taken to get out is recorded. Fifteen seconds after he starts a person is urged to go quickly. This is repeatedly used to create emotional stress. The number of times a person touches or penetrates a side line or obstacle is used to calculate his error score.

The pencil and paper form of this test has the advantage over an electrically recorded form that a complete record is retained for comparative study. It has the limitation that what the psychologist says or does is the only emotional stress which can be applied.

Gibson shows that, for normal children, T (the time taken to get out) decreases up to the age of 10 at least, but the E (errors) 'do not appear to alter significantly with age in normal populations'. He therefore calculates 'adjusted E scores' in which T is partialled out to his own satisfaction.

In some interesting tables and graphs Gibson shows for particular groups of children the time to get out, as well as the number of errors made, but it is not easy to evaluate group behaviour in terms of whether it is quick or slow as well as accurate or careless, as these two variables themselves correlate 'around −0·5'. From these tables and graphs, partialling out the time variable does not appear to be the solution.

To compare errors of behaviour under stress, time could be kept constant by using a track which became visible and moved towards a person at regular speeds. Unfortunately this requires suitable apparatus. The merit of Gibson's test is its simplicity. For social studies this is important. In clinical work it is less essential.

J. C. RAVEN

Neurological Mechanisms of Hearing and Speech in Children. By IAN G. TAYLOR. Manchester: Manchester University Press. 1964. Pp. xii + 237. 42*s*.

The functional organization of the auditory system is currently the subject of much animal experimentation. But, as vehicle for the reception of verbal messages, a uniquely human function, the complexities of the auditory system are maximal in humans. Experimental access to the human brain is not readily available, but electrical recordings may be taken from the surface of the scalp, and at times of the brain; naturally occurring disease as well as its surgical remedies may induce central auditory system changes measurable by behavioural as well as electrical means. Delayed or arrested development may give rise to 'central deafness', a conveniently abbreviated title for a rich variety of disorders of central auditory processing. Professor Taylor's mastery of this symptom complex is very evident in the monograph under review, which deals with the variety of causes of childhood deafness, with emphasis on those of central origin. The centrally deaf child, with a bewilderingly variable response to noise, but failure to learn to understand speech, mystifies both his parents and their medical advisers. Suspected of deafness, mental defect and emotional disorder, he may well acquire at least the last of these during prolonged journeying from one misconceived diagnosis and inappropriate therapy to the next. The diagnosis of central deafness is a difficult but important exercise. Prof. Taylor has made a systematic attempt to define criteria and his admittedly elaborate methods nevertheless hold promise for generally applicable diagnostic advances.

A concise summary of present knowledge of the physiology of the auditory system, and the diseases which may modify it, is followed by descriptions of the two main forms of 'objective audiometry', the sleep encephalogram, described in its detailed application to children, which is preferred to the measurement of psychogalvanic skin resistance in relation to a conditioned auditory stimulus. A series of simpler behavioural tests of auditory acuity are then described, and some useful points made in relation to parent guidance and home training. There follow relevant documented accounts of children with peripheral deafness, dysphasia, mental defect, cerebral palsy and visual handicap, followed by a section which presents the results of differential diagnosis by sleep encephalography of cases in these groups, an original contribution of major importance. A brief discussion concludes the study.

MARCEL KINSBOURNE

Speech Disorders—Aphasia, Apraxia and Agnosia. By Lord BRAIN (2nd edition). London: Butterworths. 1965. Pp. 201. 47*s.* 6*d.*

The first edition of this book introduced order and clarity into a subject noted for confusion. The present one brings the discussion and the presentation of relevant clinical material up to date.

The broad design of the book remains the same. As before an introductory chapter consists of an exemplarily detached account of theories of the nature and origins of language while the greater part of the book is devoted first to an extensive historical analysis of the growth of medical knowledge about the three conditions which form its subtitle (and related disorders such as alexia, agraphia, amusia, acalculia) and secondly to the full consideration of all the relevant clinical aspects.

A new chapter specifically treats current views of aphasia, while an expanded one on the subject of handedness and cerebral dominance in their relation to it now ends with a section on the practical and theoretical significance of the intra-carotid injection of sodium amytal.

A revised chapter deals with language and speech disorders in childhood, but the treatment remains somewhat sketchy. For example the (for the psychologist) controversial topic of specific or developmental dyslexia and associated dysgraphia is accorded about two pages. A more detailed examination of this current problem from so authoritative a neurologist as Lord Brain would be welcome. However, adequate and up-to-date references are given in this chapter (as elsewhere throughout the book with one exception), so that the reader with special interests can trace the appropriate important literature.

The inadequate bibliography referred to relates to the chapter on the development of speech in the child. A wealth of psychological literature, largely American, exists on this subject but is not mentioned.

In his theoretical approach to the neurology of speech and language Lord Brain retains the hypothesis of the schema and allies it with an adaptation of probability theory. This notion may be useful as a descriptive tool. Present limitations in the modes of neurological investigation still preclude the testing of the explanatory and predictive values of the theory.

This book remains a valuable model of interdisciplinary communication permitting language specialists in psychology and other fields to check the validity of their theories against the observations made up to the present within the framework of neurology and speech pathology.

A. W. MASON

Three Theories of Child Development—The Contributions of Erik H. Erikson, Jean Piaget, and Robert R. Sears, and Their Applications. By HENRY W. MAIER. New York: Harper and Row. 1965. Pp. xvi + 314. Standard edition 60*s.* International Students' edition 35*s.*

The linking of theory and practice in psychology presents peculiar difficulties, and any systematic attempt to overcome these deserves careful attention. This is especially true of an attempt at practical synthesis of two or more major theories. The programme of this book is, in principle, entirely commendable: it is, first, to summarize and compare the developmental theories of Erikson, Piaget and Sears; secondly to analyse the professional helping process; and finally to show how concepts drawn from the three theories can be used to guide the practitioner in thinking about his work with young people at each stage of childhood and adolescence, with their parents and with caretaking agencies.

Admirable scheme; but does it work? The author's craving for system (in a field where system is admittedly needed) leads him at times to compare the incomparable. While the distinctions between affect, cognition and behaviour are by no means absolute, it remains true that the work of the three writers in question is centred in these respective areas to a degree that makes their output complementary rather than competitive. Although Maier acknowledges and even stresses this (p. 206), he nevertheless becomes involved in teasing out tenuous and often confusing similarities and differences of view. Most disappointingly, on the other hand, he neglects to elaborate (except in relation to therapy in his final chapter) the integrated scheme he charts on page 189, based on phases of dependence–independence, which appears to hold possibilities of at least a partial synthesis, and merits exposition in its own right.

It is nowhere stated that Erikson, Piaget and Sears gave their approval to the summaries of their thought offered here—a sanction we may feel entitled to demand when the theories of still active writers are presented at second hand. While the general presentation seems fair, if not always clear, and documentation is good, occasional statements are highly questionable. Thus, he states (p. 161): 'Sears implies that the commonly assumed socioeconomic class differences between persons do not exist.' What about chapter 12 in *Patterns of Child Rearing*, which is devoted entirely to the bearing of such differences on the treatment of children?

Perhaps the most useful chapter contains the author's analysis of the helping process. He sees diagnosis and treatment as 'associated and continuously interwoven processes' and analyses the various interpersonal systems through which help can be given. Regrettably, this chapter, like the rest of the book, is couched almost entirely in abstract terms. Concrete examples are sparsely used, seldom running to more than a sentence, and never personalized. In a book by a clinician addressed to clinical workers, is it not strange to find never a single case study, however brief, not one real child to bring the theories to life? It seems odd that a professor of social work should show so little empathy toward his readers as to offer them this incorporeal diet of abstractions.

Much thought has certainly gone into this book and it must not be dismissed out of hand merely because the reviewer happens to dislike verbiage. Any clinician feeling the need to structure his thinking about child development and about his own role is advised to sample for himself what Maier offers. If he is prepared to translate a bookful of general statements into the particulars of his own experience, he may well find the theoretical schema helpful.

TERENCE MOORE

The Psychology of Childhood and Adolescence. By C. I. SANDSTRÖM. London: Methuen. 1966. Pp. 260. 36*s*.

Apparently this book has not only been very successful as a text for university students and teachers in training colleges in Sweden but 'has also been something of a "best seller" to the general reader and to young parents who take an intelligent interest in the development of their children'. However, there are several reasons why one would hesitate to recommend it as a text-book in this country. The author's aim to 'write a survey not exceeding about 250 pages' has led him to include too many topics in far too condensed a form: a historical survey of developmental and educational psychology, covering the past 200 years, is presented in ten pages; 'Motivation and the development of learning' are covered in twelve pages; 'the aims of factor analysis' are described in one page and the effect of 'incomplete homes' in under half a page.

Its value as an introductory text is limited by the fact that references are very inadequate: only author's names are cited, and the full reference is rarely given. At the end of the book there is merely one page of 'Suggestions for further reading'.

It is not always apparent whether a certain naïveté and lack of clarity are due primarily to infelicitous translation or whether they are a reflexion of the author's thoughts. To give some examples: 'When we say that there is joy in work in a class, we can usually also say that the children are well motivated in relation to the school'; or 'It is a question of maintaining interest in definite perception situations so that learning will be possible.'

Though Sandström warns from time to time against too ready generalizations about children's development, his outline of growth stages and norms suggests a much greater uniformity than in fact exists. For this reason alone, one would be reluctant to recommend the book to parents.

A really suitable introductory text-book on the psychology of childhood and adolescence for students in colleges of education and university departments of education and psychology remains to be written. In the meantime, however, this book is unlikely to take the place of the many American texts which have a similar aim; though generally much longer, they do cover the field more adequately as well as more readably.

M. L. KELMER PRINGLE

Handbook of Clinical Psychology. Edited by B. WOLMAN. New York: McGraw-Hill. 1965. Pp. xv+1596. 196*s*.

This is a work containing fifty-eight chapters grouped into six sections. A handbook of this size is presumably intended to give a full coverage of the field, although not in the detail expected from a less wide-ranging work. By the definition of the *O.E.D.* it should contain 'concise information' which will enable the reader to gain an adequate picture of that aspect of the subject on which he requires information, together with the necessary references to widen his range of reading. It is also to be hoped that, while some authors may be permitted to grind particular axes, the overall choice of contributors will be such as to give a fully rounded picture of the field.

The editor's stated purpose 'is to acquaint clinical psychologists and other professionals with the tremendous scope of research, experience, theory, and practice in this rapidly growing field, ... and to demonstrate its vitality, its vigorous pursuit of scientific truth and its willingness and capacity to serve those who need it'.

It is against these aims that the book may be judged.

The contents of the first part, entitled 'Research methods', are perhaps best described as the application of scientific method in clinical research. In general the contributions are all worth while, and particularly so is the chapter on statistical issues, which tackles, amongst other things, the use of one-tailed tests and non-parametric statistics with considerable clarity and wit.

The contributions to the second section, which deals with 'theoretical foundations', are for the most part excellent and up to date although a British audience might wish for less than 70 pages to be devoted to analytic theories. There is inadequate coverage of recent experimental psychological contributions to the understanding of the psychoses.

Part 3, which is concerned with diagnostic methods, seems least to exemplify the editors' claim to demonstrate vitality and vigorous pursuit of scientific truth. No attempt is made to provide a 'forward look' which might put into practice the theoretical positions outlined in section two, and the reader is condemned to nearly 200 pages on projective tests. The remainder of this section is mainly devoted to the interview, personality inventories and the diagnostic use of intelligence tests.

Part 4, the longest in the book, is concerned with clinical patterns all of which are reasonably well covered, although again the British reader might justifiably complain of an excessive analytic bias. It is a matter of some interest that this section contains a long chapter by Luria on focal brain lesions.

Part 5 on methods of treatment might be thought to be of least interest to British readers in view of the differences in climate of opinion between this country and America on the use of treatment methods by clinical psychologists. Apart from the coverage of psychotherapeutic methods, the section contains good chapters on behaviour therapy and psychopharmacology, together with a chapter on the role of the mental hospital.

The final section is concerned with training in clinical psychology, its relation to other professions and an interesting review of the position of the discipline throughout the world.

In summary, this is a book which most clinical psychologists would most probably wish to see in their libraries, as its cost is high for personal purchase. As a reference work it is patchy; some chapters provide good coverage with extensive and useful bibliographies, others omit important references.

It appears to reflect both the conservative attitude attributed to many clinical psychologists in America and the more forward-looking attitude of experimentalists in this field. It seems a pity that, in a work which, because of its size, will not be repeated for several years, a greater attempt was not made to present more possibilities of advance by the integration of theoretical findings with practice.

P. H. VENABLES

Mental Health on a New Housing Estate. By E. H. HARE and G. K. SHAW. London: Oxford University Press. Maudsley Monograph No. 12. 1965. Pp. 135. 38*s*.

This book reports the methods and findings of an investigation carried out with the intention of comparing the mental health of people living in a new housing estate on the edge of Croydon with that of the inhabitants of an old-established and densely populated central area of the same borough. The comparison was thorough and systematic; information was drawn from mental hospital records of admissions and outpatient attendances, from general practitioner consultation records and from household interviews. In respect of virtually every index of mental health the two populations proved to be indistinguishable, and the authors conclude that rehousing either has no effect on mental health or, if it has, that this is counterbalanced by other factors. The authors see their findings as a refutation of an hypothesis tentatively advanced in an earlier investigation, to the effect that certain types of rehousing programme might have adverse consequences for community mental health; equally, however, the Croydon results disappoint the expectations of the many reformers who have predicted improvements in the mental as well as the physical health of rehoused populations.

If this study has a fault, it lies neither in the detailed planning and execution of the research nor in the reporting, which were obviously meticulous, but in the inadequacy of the basic conceptualization. The epidemiologist who embarks on a comparison of community levels of mental health (or rather, ill-health) should surely be armed with reasonably sophisticated sociological hypotheses concerning the community characteristics presumed to be relevant and select areas in which these characteristics can be studied in depth. In the Croydon study, the point at issue was *not* whether a 'lack of social amenities' somehow produced a higher frequency of mental ill-health, but whether either the short-term or the long-term modifications of patterns of personal and family life that are entailed by particular types of rehousing programme may unduly often precipitate minor psychoneurotic and psychosomatic disturbances. This notion needs to be refined and elaborated; and, if it is to be tested, the sociological aspects of the investigation require no less attention than the statistical and psychiatric aspects. To neglect these considerations is to come perilously close to comparing people who live in new houses with those who inhabit older ones. And it would be quite remarkable if such a comparison proved to be illuminating.

F. M. MARTIN

Studies on Psychosis. Descriptive, Psychoanalytic and Psychological Aspects. By THOMAS FREEMAN, JOHN L. CAMERON and ANDREW MCGHIE. London: Tavistock. 1965. Pp. vii + 245. 35*s*.

In this book two psychoanalytically trained psychiatrists and a clinical psychologist successfully demonstrate how the descriptive, psychoanalytic and experimental psychological approaches complement each other. Psychoanalysis has contributed a great deal to the study of the psychoses. Bleuler's notions of the hierarchy of symptoms in schizophrenia owed much to Freud. The psychoanalytic approach has had a large share in the change of attitude to the schizophrenic, who has proved to be much more amenable to interpersonal relationships than had previously been believed to be possible. The purpose of psychotherapy is to help these patients to re-experience past conflicts.

The language of psychoanalysis has been helpful in the description and interpretation of psychotic symptoms which are viewed as regressive manifestations. However, the question why regression occurs is still unsolved. Freeman, who is the main author, gives an excellent account of the psychoanalytic views on the major psychoses. He does not exclude the possibility of primary organic factors being responsible for the impairment of the ego functions in schizophrenia. Hallucinations and delusions are regarded as secondary phenomena serving the maintenance of relations to the outside and inner world. The chapters dealing with the patients' reactions to experimental situations and with the effects of distractive stimuli on schizophrenic performance are of special interest to experimental psychologists. Schizophrenics were found to be less able than normal subjects to improve their performance by utilizing increasing degrees of contextual organization. The normal processes of selection and inhibition had broken down. These findings may prove to be relevant to the therapeutic handling of such patients. The last chapter deals with the psychoanalytic treatment of the psychoses at Chestnut Lodge, a well-known American private hospital which specializes in this method.

Nobody who enters this area can expect clear views. Every guide light is to be welcomed, however dim. The authors have given an honest account of what psychoanalysis has so far contributed in this field. How much or how little it is worth only the future will show. This book deserves to be read by all who are interested in mental illness whatever their theoretical approach may be. Clinical psychologists in particular will benefit from being introduced to an approach which will help them in understanding not only their patients but also some important aspects of present-day psychiatry.

E. STENGEL

Patterns of Meaning in Psychiatric Patients. By I. M. MARKS. London: Oxford University Press. 1965. Pp. xvi + 144. 55*s.*

The author gave Osgood's Semantic Differential to groups of normals, obsessionals and psychopaths. Known statistical probabilities reduce the ambiguity of the statement that the groups were half male and half female. Three personal concepts (myself, my father, my mother) and three emotional areas (anger–hostility, fear–anxiety, love–affection) were utilized. The aim was to map out the semantic geography of these areas of conceptual meaning in psychiatric patients.

The psychological aspects of the study were carried out with great thoroughness. These included short- and long-term stability of the measures, their validity and interrelationships. The statistical treatment is exhaustive and the level of argumentation in the discussion and interpretation of the results is high.

The usual general evaluative factor was found and, in controls and obsessives only, a subsidiary evaluative (danger or risk) factor. Personal concepts differentiated between groups better than did emotional concepts. Psychopaths and obsessives used more extreme scale positions than did controls and were again contrasted with controls in having a disturbed self-concept; but psychopaths alone had disturbed parental concepts. Psychopaths again stood out from the other two groups in having rather more fear of anger–hostility, whereas, on anxiety, psychopaths and obsessives contrasted with controls.

The psychiatric side of the study is less satisfactory. The obsessives were selected on the basis of the usual symptoms of obsessive–compulsive neurosis. Fifteen of the twenty showed some obsessive traits. It is surprising that all twenty did not show some obsessive traits in some degree. What differentiates people, surely, is the preponderance or strength of particular types of traits rather than their presence or complete absence. It is quite conceivable that the psychopaths had obsessive traits. Were they looked for? It is particularly possible when the criteria for inclusion in the psychopathic group were: repeated physical violence; repeated shouting, screaming, general abusiveness; repeated theft, robbery, larceny or fraud; drug addiction. Persistent thieves can probably pick out psychopaths from amongst themselves better than that, even though they may not use the term. These inadequate criteria weaken the otherwise justifiable and important conclusion that the results support the clinical view that psychopaths in psychiatric institutions have feelings of guilt.

Finally, psychopaths and obsessionals seem to have been chosen because, within psychiatric groups, they represented polar opposites in respect to impulsiveness. What then is the normal control group controlling? It is not necessary. Having got it, we could say that in respect to impulsiveness—or, more broadly, extraversion—the controls should fall mid-way between psychopaths and obsessives (P–C–O). In terms of such deviance as brings one to a psychiatric institution (for short illness), psychopaths and obsessionals should contrast with controls (PO–C). Personality–illness interaction should produce something like P - - - O–C, which, in most instances, is what is found. The results could have been more unequivocally explained had the samples been, say, extraverted and introverted psychopaths, obsessionals and normals, though hysterics might have been a better choice than psychopaths.

This valuable investigation should encourage further work with this technique; but it is to be hoped that it will be published at a less exorbitant price.

G. A. FOULDS

Pavlovian Psychiatry: A New Synthesis. By CHRISTIAN ASTRUP. Springfield, Illinois: Charles C. Thomas. 1965. Pp. xiii + 171. $6.75.

Dr Astrup is a Norwegian psychiatrist of international distinction. He is perhaps best known in this country for his work with schizophrenics, using Leonhard's classificatory method, and is thus often taken to be a Kleistian. In fact, he has had wide experience of conditioning procedures and is pre-eminently qualified to interpret and assess Pavlovian developments in psychiatry from several points of view.

This is physically a very pleasing book. Clearly a great deal of care has been taken in its production and editing. In surveys of this kind it is difficult to avoid stodgy reiteration and a jerky rhythm; Dr Astrup's style is relaxed, lucid and readable throughout. He manages to spare the reader that bemusement which often accompanies the ingestion of terminology translated from the Russian.

The importance of this book lies in its survey of research reports not available in English. The reference list contains 443 titles. Out of these about 260 are from the Russian literature, and many others are from Scandinavian, Polish, East German and other sources. In his preface Dr Astrup claims that the reference list represents merely a supplement to the bibliography of his *Schizophrenia: Conditional Reflex Studies* (1962) and that in the present work he has 'paid special attention to recent research'. In fact, the few post-1962 references cited here are almost exclusively papers read to the 1963 International Congress of Psychiatry and the 1964 International Congress of Social Psychiatry.

A further limitation will immediately strike psychologists who read this work: that is, the paucity of references to non-medical researchers. A few standard psychological texts—Hilgard and Marquis, Hull, Mowrer, Skinner, Watson, Woodworth and Schlosberg—do indeed appear in the reference list. But only Skinner and Watson rate mentions by name in the text. Nor is this restricted to non-Russian psychologists. Luria receives a passing mention in the text but nothing in the reference list. Teplov, Beritashvili and Myasishchev are among those disregarded. Again, the flourishing field of behaviour therapy receives only the sketchiest consideration. In the relevant chapter Wolpe is given a couple of pages, but these refer only to his 1958 book on reciprocal inhibition. Of the few other workers mentioned in this section the only English-language writers are, perhaps surprisingly to some, Slater and Sargant. No attention whatsoever is paid to more sophisticated developments in this field.

Of particular interest to psychologists are the sections dealing with Pavlovian approaches to the functional psychoses. Whether these will have a ready appeal to clinical psychiatrists is a different question.

In short, clinical psychologists will find this book valuable for its Russian references; its textual content they will find interesting but unsatisfying. Learning and personality theorists will probably be disappointed by its conceptual level, which is much below that of Gray's recent contribution to Pavlovian typology. As an introductory text for psychiatrists, however, this could be admirable. Certainly every D.P.M. student should be encouraged to read it.

GRAHAM REED

Hypnotic Susceptibility. By E. R. Hilgard. New York: Harcourt Brace and World. 1965. Pp. xiii + 434. $9.50.

The Nature of Hypnosis. Edited by R. E. Shor and M. T. Orne. New York: Holt Rinehart and Winston. 1965. Pp. xii + 504. $10.95.

Books on hypnosis continue to appear at intervals regular enough to suggest that there is a ready sale for them, even if their contents confirm that developments in this field are failing to keep pace with the rate at which the print pours off the presses. While neither of the present titles is likely to become a classic, either on account of epoch-making revelations or deathless prose, they are both perfectly reasonable and readable, and may be recommended to anyone who wants an up-to-date exposition on the topic—until the next perfectly reasonable and readable text plops just that little bit hotter (and wetter) off the press.

In Hilgard's opus there is plenty of fodder for those who enjoy ruminating over lush pastures of correlation matrices. The upshot of all this exercise in free-range animal husbandry is that there is a general factor to be discerned, plus a larger indefinite number of other factors associated with susceptibility to hypnosis. The susceptibility scales themselves will doubtless be of value to anyone engaged in this particular research area. Much effort has gone into compiling and evaluating them, and the reader is spared very little of it.

Towards the end, when scientific endeavour has given way to intuitive speculation, a spark of suppressed originality may be detected. At this stage some interesting ideas emerge, such as attention being drawn to the parallel between becoming immersed in a good book or film and the dissociation characteristic of hypnosis. Nor is it unreasonable to suggest that the ability to be drawn into and abandon oneself to the world of fantasy is related to hypnotic susceptibility. Other personality variables are sniffed over and rummaged through, including the traits of mental and physical adventurousness.

The second book, *The Nature of Hypnosis*, is rather more original, being a collection of previously published material from a variety of sources. A good deal of innocent fun is to be derived from some of this early literature, much of it being concerned with questions of moral welfare. It is easy to see now why Mesmer's technique was more effective with women than with men, at least in his hands. There are several other fascinating snippets, including Galton's clipped account of how he drove himself mad in the cause of science.

The controversy between the sceptical and credulous views of hypnosis makes pretty dismal reading, in that a lot of misplaced effort seems to have gone into casting doubt on the 'genuine' nature of hypnotic phenomena, although no one quite has the nerve to suggest that Esdaile's patients were simulating lack of pain as their limbs were amputated or eyes plucked out without the benefit of chemical anaesthetics.

Signs of a *rapprochement* are to be found in Shor's contributions, where a distinction is made between withdrawal from reality (trance) and role-responding to the 'demand' characteristics of the situation (hypnosis). Hypnosis is thus regarded as a form of trance development peculiar to a particular culture pattern. Degree of archaic involvement (regression) is offered as a third dimension, and presumably the concept of dissociation will have to be accommodated somewhere along the line, unless it is intended to be subsumed already under the trance rubric. Thrown in for ballast is a Cook's tour of some of the current theories, with the notable exclusion of more modern physiological theories involving the reticular activating system.

In short, if you have not read a good book on all the things that have not been happening in hypnosis over the last ten years, now is your chance to catch up. But do not worry if you miss either of these two buses—two more with the same numbers on the front are just coming round the corner for sure. On this route there is a good service.

R. T. GREEN

Psychology in Industry. By NORMAN R. F. MAIER (3rd edition). London: Harrap. 1965. Pp. xiv+718. 50*s*.

The first edition of this book came out in 1946 and the second edition 16 years later. The second edition was not very different from the first, and some important libraries did not see the need to purchase it. The present edition is, however, a substantial development from the first.

The first notable addition is that at the end of every chapter there is a 'laboratory exercise' in which students are given material for role-playing a situation which relates to the content of the chapter.

Practically every chapter has been expanded, and modified by substituting more recent illustrative examples for the older ones. This generalization applies not only to the text, but to the layout of paragraphs and tables, the improvement of figures and diagrams and even the typeface of captions, headings and explanatory text.

After recently seeing the readiness of psychologists overseas to purchase American tests and assume that they are applicable in their own country, the reviewer was delighted to read Maier's warning on the misuse of tests. The chapter entitled 'Motion and Time Analysis' has been completely revised and rewritten by Hoffman of Michigan. A careful and interesting section sets out the criticisms and limitations of the former approach and leads in to the modern biopsychological approach to design of jobs. The phrase 'human factors' is used in preference to 'human engineering' but the words ergonomist and ergonomics are only just mentioned and discarded.

The extent to which development has taken place in this area is to be measured not only by the dropping of so much of the text of the older edition but by the very substantial new material that is included. Even so it might be felt that this chapter does not do full justice to some of the most recent work. Although the section on training is also greatly enlarged and touches on most of the up-to-date issues, in some ways there is a feeling of disappointment in this, as in some other chapters. To take an example, the reviewer tried to find what validation there is on T-Group training. The T-Groups are dealt with, but only in a few short paragraphs, and in this instance even the references do not satisfy.

Perhaps the time has come when the general book on the whole of industrial psychology cannot but fail to be totally satisfactory. The field has grown so much that a single book cannot do more than skim the surface. But Maier has done better than that. If the book is intended less for the professional psychologist than for the industrial executive or personnel manager he has been outstandingly successful. Even the professional industrial psychologist must be grateful, for the book covers a very wide range and is generally very adequately up to date.

There is no doubt at all that your reviewer will have constant use of this valuable book even though it will still be necessary to go to the journals for anything much more recent than 1960.

H. G. MAULE

Human Engineering Guide for Equipment Designers. By Wesley E. Woodson and Donald W. Conover (2nd ed.). Los Angeles: University of California Press. 1964. Pp. 470. $10.00.

The first edition of this book was published in 1954 as a practical guide to the design of controls, machines, furniture and buildings in the light of knowledge about human physiology, anatomy and psychology. Although it was criticized for its cartoon strip approach, its tendency to oversimplify and its somewhat arbitrary recommendations, it came to be widely used, and was translated into many languages. This second edition is an enlarged and updated version of the first, still written in a direct, self-confident style and humorously illustrated. The additions are mainly in the introduction, where sections on design philosophy, cybernetics and bionics have been included; and in the chapter on body measurements.

The authors, both of whom work on problems in air and space flight, clearly intended that their book should educate the engineers in the human factors of equipment design rather than just be a reference guide. This intention has been emphasized and developed in the second edition. However, many readers will still be disturbed by the lack of references to source data in the text and the absence of qualifying clauses to more controversial recommendations. No means and standard deviations are reported, nor are the type and size of the measured populations stated. The book sets out to be comprehensive as well as precise, but when recommendations are made for the design of ironing boards and bedroom dressing tables further doubts are encouraged about the sources of the authors' data, and consequently the validity of their recommendations. Is their source 'common sense' or are they generalizing from samples of one? Checking through all the authors' recommended designs for working chairs and seats it becomes apparent that they settle on the 18-inch seat-surface to foot-rest distance. This is the 'traditional' seat height, and obviously upheld by the statement in the text 'seat heights are expected to be at a certain level when a person sits down'. However, Swedish and British work points to between 16 and 17 inches being more the optimum seat height for adult males, and women in high-heeled shoes.

As a summary of the considerable amount of experimental and field work on the optimum measurements for the design of equipment for human working use this book is inadequate. As a spur to design engineers to take into account some of the basic principles and parameters of human performance this second edition can be recommended.

G. A. RANDELL

Human Behavior: A Program for Self-Instruction. By Leslie F. Malpass (editor), Max D. Hocutt, Edwin P. Martin and Paul R. Givens. New York: McGraw-Hill. 1965. Pp. xvi + 400. $7.50.

This is an introductory text-book in the form of a linear program. The four authors contribute six sections. Hocutt gives an elementary but useful introduction to 'the methods of behavioral science'. The second section by Martin on the 'biological bases of behavior' begins with homeostasis, going on to simple genetics and having a quick stab at the nervous system, the receptors and the endocrine glands. In Part 3 Malpass introduces 'principles of learning behavior', beginning with a short piece on perception and followed by brief accounts of classical and operant conditioning. Part 4, also by Malpass, takes us from conception to marriage (somewhat after the manner of Dr Spock). Part 5 by Givens introduces personality theory, principally Freud, Murray and Roger. Part 6 sees the return of Hocutt with a short section on 'mind, body and behavior' which attempts to define a few of the psychological 'isms'.

The authors apologise for superficiality, as well they might, and there is a wealth of those quasi-truths one finds in first-year essays. My favourites from this book are, 'College Sophomores have been conditioned to blink their eyes at many things...', and a photograph of a smashed-up car with the legend—'A possible result of unconscious reactions and impulses'.

The book should also be considered for its qualities as a program. Each section begins with a prose summary and continues as a linear constructed-response program in vertical format with the answers in the margin. At the end of each section is a self-test quiz and a bibliography. Inside the cover is a table cross-referencing a dozen texts with sections of the program. A program can really only be judged by its results but in the present case objectives are not closely specified and no results of evaluation studies are quoted. If psychologists do not follow the APA recommendations on evaluation how can one expect anyone else to do so? The programming style rarely rises above the crude 'Swiss cheese' variety and one feels the authors, and perhaps the readers, would be happier with straightforward prose. Skinner's techniques have been imitated rather than used. For one who is still hopeful of the possibilities of programmed learning in higher education this book comes as a disappointment. A second volume on social psychology is promised.

JOHN ANNETT

The Structure of Behaviour. By M. MERLEAU-PONTY. Trans. A. L. FISHER. London: Methuen. 1965. Pp. 253. 35*s*.

This book is a work in philosophical psychology. It is an attempt to defend a holistic and structural account of behaviour by a detailed consideration of recent work in psychology and physiology. It springs from the philosophical tradition of phenomenology, especially the work of Husserl, which means that the average British reader will find much of it very difficult to follow. It exerts, however, a queer sort of fascination, because in the fog in which the reader so often finds himself he comes across bright patches that encourage him to persevere with his journey.

Chapter 1 is on reflex behaviour. It argues, against reflex-arc analysis, that response is determined by the total pattern of the situation and by internal factors of an organic and psychological nature. Some sympathy is shown towards gestalt theory, but it is criticized for its tendency to project human norms into phenomena and to suppose 'directed' or 'ordered' processes.

Chapter 2 deals with higher forms of behaviour. Pavlov's reflexology is criticized for extending the term 'conditioned reflex', which should designate a relatively stable reaction linked with certain stimuli, to cover reactions that are variable and can be dissociated or even reversed. Merleau-Ponty then passes to work on 'the central sector' (e.g. by Gelb, Goldstein, Head and Lashley). He objects to theories which think it sufficient to correct atomism by notions such as those of co-ordination and integration, and argues that the physiology of the brain cannot be properly conceptualized without borrowing from psychology. He then begins to elaborate his notion of 'structures of behaviour'. He rejects behaviourist and gestalt theories of learning and maintains that, in learning, 'the decisive factor lies in the manner in which fortuitous contiguities are utilized by the organism, in the elaboration which the organism makes them undergo' (p. 100). Trials themselves never contribute to the acquisition of a habit. They must occur in a certain organic framework which gives them meaning and efficacy; they presuppose a 'sensory-motor *a priori*,' practical 'categories' which differ from one species to another.

'Syncretic', 'amovable', and 'symbolic' forms of behaviour are then distinguished and discussed. At the 'syncretic' level behaviour is tied either to certain abstract aspects of the situation or to complexes of special stimuli. It is imprisoned in the framework of its natural conditions. As soon, however, as *signals* appear which are not determined by the instinctual equipment of the animal, we can presume that they are founded on structures which are relatively independent of the materials in which they are realized (Tolman's notion of 'sign-gestalt'). The notion of taking a means to an end begins to have application. Finally, there are symbolic forms of behaviour in which signs function as symbols, not just as signals. The sign ceases to be merely an event and becomes the proper theme of an activity which tends to express it—as with an organist playing a tune from a score.

Chapter 3 is on the physical order; the vital order; the human order. It is an attempt to understand matter, life and mind as three orders of signification, characterized by different types of structures. What we call a physical system is an *ensemble* of forces in a state of equilibrium or of constant change such that each vector is determined in size and direction by all the others. But structure and law are 'two dialectical moments and not two powers of being' (p. 142); for ultimately form can only be defined in terms of knowledge. 'Laws have meaning only as means of conceptualizing the perceived world' (p. 145).

We speak of 'vital structures' when equilibrium is obtained, not with respect to real and present conditions but with respect to conditions which are only virtual and which the system itself brings into existence. Each organism, in the presence of a given milieu, has its optimal conditions of activity and its proper manner of realizing equilibrium determined by its general attitude towards the world. Thus organic structures are understood only by a norm, not expressed by a law. Of particular importance are the organism's thresholds of perception, which, linked with its preferred attitude, give a new kind of unity to behaviour.

The human level is characterized by 'use objects' (e.g. clothing, tables, gardens) and 'cultural objects' (e.g. books, language, musical instruments), which constitute the proper milieu of man and bring about the emergence of new types of behaviour. 'Work', the activity by which man transforms physical and living nature, is more distinctive of man than behaviour. 'What defines man is not the capacity to create a second nature—economic, social or cultural—beyond biological nature; it is rather the capacity of going beyond created structures in order to create others' (p. 175).

In the first three chapters the general drift of Merleau-Ponty's criticisms of psychological theories can be gleaned; the general shape of the conceptual distinctions which he thinks crucial in theorizing about human behaviour are readily discernible. Much is said that is acute and suggestive and it is indeed welcome to find a philosopher writing about psychological theories with such a detailed grasp of their content. When, however, the reader embarks on the fourth and final chapter, dealing with the more purely philosophical topic of 'The relations of the soul and the body and the problem of perceptual consciousness', he may well find himself completely befogged. It is not altogether clear, to put it mildly, what Merleau-Ponty considers the problem of the soul's relationship of the body to be. His discussion, which stresses the importance of perception and of structure in this relationship, is very difficult to follow. His main theme is that experience is our way of being inextricably involved with the world through our felt body.There is an affinity, at different levels of structure, between consciousness and nature. This sounds acceptable enough as a general thesis; but its particularities are very perplexing.

R. S. PETERS

Language, Thought, and Comprehension: A Case Study of the Writings of I. A. Richards. By W. H. N. Hotopf. London: Routledge and Kegan Paul. 1965. Pp. x+349. 40*s*.

The sophisticated but unwary psychologist who opens this book, having read its title but neglected its subtitle, is in for a sharp shock, which will be all the more catastrophic if he is young enough to have a yawning historical gap between the dim wraiths of Würzburg and the contemporary psycho-linguists—Skinner, Osgood, G. A. Miller, and the rest—who go unmentioned here, along with all the fashionable apparatus of plans and the structure of behaviour, information theory, cybernetics and Markov chains. Even this reviewer was shaken to realize that the resounding splash occasioned by *The Meaning of Meaning*, from which this book starts, occurred almost half a century ago.

An astonishing book it is—difficult to write, difficult enough to read and harder still to review, especially for one who would not claim the polymathic erudition of its author, and who is painfully conscious of the warning, on p. 144, of the risks that lie across the would-be critic's path: 'Who has written a careful book on (for instance) "economics, philosophy, criticism, social theory" without having privately to admit that all his reviewers have misread him?' But then, 'Language and the Diseases of Communication' was one of the Wagnerian *motifs* of Richards' work, and Mr Hotopf has already contributed his own variations on this theme in his earlier entertaining explorations of the correspondence columns of the newspapers, in the form of some psychological *obiter dicta* that gave no hint of the present *tour de force*. An astonishing book—if indeed, it is not rather three books in one, a three-decker case study that is the outcome of years of reading of Richards, the most significant English critical writer of the present century, who contributed not to linguistic and literary theory only, but also to educational theory and practice, and to the movement of psychological and philosophical thought between the wars.

No one would dispute Hotopf's claim that 'Richards is worth a book', or that, to be satisfactory, that book would have to be, as Hotopf's is, interdisciplinary in its approach. Nor, I think, would it be disputed that Hotopf was unusually well qualified to undertake the task as a social psychologist with a particular interest in problems of communication and an initial training as a philosopher and historian. The psychological reader cannot expect, therefore, that Hotopf's discussion should be narrowly psychological, or, in the commonly accepted sense, strictly technical in its method. Nor, if he is to read the book as a contribution to a field too cavalierly neglected by the professional psychologist, can he afford to disregard the other aspects of Richards himself, or of Hotopf's discussion. Hotopf, indeed, indicates that he has been particularly concerned with assessing Richards' wholesale application of psychological explanations within the different disciplines.

The Meaning of Meaning apart, Hotopf confines his study to five of Richards' main books. The discussion then falls into three main sections as follows: (i) a critical analysis of the content of these publications (chaps. 1–5); (ii) a critique, on the basis of Richards' own methods, of the 'misunderstandings' and 'failures of communication' that have been demonstrated by a number of the most significant philosophical and literary commentators on Richards' work (chaps. 6 and 7); (iii) the conclusion of the whole matter, and certainly the most interesting section of the whole as representing Hotopf's own reasoned evaluation of the Richards material, and in particular of the lines that he had taken concerning the meaning of meaning, language, language and communication, and language and thinking (chaps. 8–11).

There is no doubt that Hotopf's book is an important one. It is concerned, as Richards was, particularly with those aspects of language—content and meaning—that have been treated most gingerly by linguists and psychologists alike in recent years. Indeed, in so far as 'unobservables' appear to be involved, the concept of meaning is one which some psychologists at least would prefer to avoid, or even to abandon. It is impossible to do justice here to Hotopf's long argument. It is perhaps a pity that Hotopf gives so much time to the critics of Richards, although the material is not irrelevant to his 'case study'; and, although I suspect that Hotopf himself enjoyed doing it, I doubt whether many of his readers will find his somewhat apocalyptic 'exercise in Freudian phenomenology' in the last chapter ('Genesis and Revelation II') any more satisfying than this reviewer did. But two points should be made. Hotopf's book is no mere exercise in memorial homage, although it is that as well. Few psycho-literary 'clinicians' (except perhaps Livingston Lowes) can have paid their chosen 'case' the tribute of such an effort at understanding. Finally, on his employment of the 'case study', Hotopf may speak for himself. In so describing the book he writes: 'I hope to suggest that the clinical method may be applied in an area where the experimental approach has so far ruled'; and while admitting the inevitably speculative aspects of the study he comments further: 'The functions of a psychologist are not limited to marching in the scientific procession towards further certain knowledge. Man is an animal who theorises psychologically, and the specialist training of a psychologist in a powerful empirical tradition ought to be able to sharpen and further articulate the experiences from which these inevitable theories come.'

GEORGE SETH

Memory. By BRIAN SMITH. London: Allen and Unwin. 1966. Pp. 214. 40*s*.

Dr Smith's book is a contribution to the philosophical discussion of memory, rather than the psychological investigation of memory phenomena. But it may be of some interest to psychologists, and may suggest possibilities for empirical investigation. As a philosophical work it is unusual in two ways. In the first place, Dr Smith does not address himself at any length to the usual epistemological problems which are raised by philosophical scepticism about memory as a form or source of knowledge. Secondly, his approach is one which will strike many philosophers as curiously old-fashioned; he relies heavily on the evidence of introspection; and he speculates about possible physiological causes of mental phenomena in a manner more reminiscent of William James and Bertrand Russell than of more recent philosophy.

The question to which Dr Smith's book is devoted is this: Given that we do, in fact, rely on our memory of past events in almost all cases, why do we do so? It is plain that between our witnessing some event which we later claim to remember and the occasion of our making the claim there is ample room for mistakes to creep in. Usually mistakes are not made, and usually we are able to ensure that our claims are justified, or that, where they are not, we are able to detect the mistake that has occurred. The problem is to explain how we do this. The half-heartedly behaviourist view, which reduces memory to our currently making claims about appropriate pieces of the past, is quite rightly rejected as indaequate. Dr Smith distinguishes between the memory-claim that such and such an event occurred and the actual memory on the strength of which the claim is made. The actual memory then remains to be characterized. It is here that many readers will find Dr Smith's account hard to follow. The reason is not the complexity or obscurity of his basic views, but the number of incidental discussions which are not clearly linked to the main theme. For example, there is a running argument to support the view that facts cannot be remembered but only asserted in propositions; propositions can be remembered, but by non-verbal means as well as verbal ones. To many philosophers these will seem perverse views to hold, and the difficulty here is that they are not sufficiently discussed to become important in themselves, nor is their place in the overall picture of what it is to remember something made clear.

The gist of the book's answer to Dr Smith's question is, I think, as follows. When we perceive events, or entertain propositions, we sense appearances, in a use of the term 'appearances' which implies only that we are sensorily stimulated in a given way. This given way may be described physiologically, and is what Smith calls a 'sensation'; and in perception we become aware of our sensations. Memory, then, involves the literal re-presentation of the appearance, in other words a repetition of the physiological process of sensation in the absence of an external stimulus. In the same way that perception is our being aware of the sensation, memory is our being aware of its repetition. And on the strength of this account Dr Smith is able to explain where we can go astray in making false memory-claims, and where we can have accurate memories, even though we may make false claims on the strength of them.

On the level of ordinary common sense, much of the book is sound reporting, and the physiological speculation is as interesting as such speculation usually is. But it does not seem to me to advance our understanding of memory very far. As a subject for philosophy, memory is not enough to sustain a book of over 200 pages; and the mixture of introspective evidence, speculation and linguistic *fiat* is philosophically disquieting. The dangers involved in talking about 'images' and 'appearances' are neither explained nor avoided, and Dr Smith is none too safe in the area of such ancient pitfalls as the notion of 'direct perception'. As psychology, I cannot believe that the book serves any very useful purpose. Memory does not seem to be one of those areas where there is great conceptual confusion, and, if it were, there is no evidence that Dr Smith is much interested in the conceptual problems of empirical psychology. Dr Smith seems uninterested in the enormous difficulties of describing the physiological mechanisms of the remembering process. He espouses a simple causal model which has seemed intuitively clear to 300 years of philosophers and which seems to appeal much less to physiologists today. Philosophers, I think, will profit from some of the discussions in this book, though hardly from the whole of it; but I should be unhappy to offer it to a psychologist as a sample of what philosophy can do for psychological questions.

ALAN RYAN

Signal Detection and Recognition by Human Observers. By J. S. Swets. London: Wiley. 1964. Pp. xi + 702. 132*s.*

This book is a collection of papers concerned with applications of the theory of signal detection first expounded by Peterson, Birdsall and Fox (1954). They are, necessarily, but a small proportion of the total number of papers which have been inspired by this work, but they are a well-selected sample and their collection in one volume will be of considerable value for tutorial and reference purposes. The only major omission seems to be the original seminal paper by Peterson, Birdsall and Fox, which until very recently was available only as an excessively long and cumbersome, privately circulated report. This report is obviously well known and used by most of the authors whose work is collected here, and it would have been of considerable help to those outside the charmed circle to have included a shortened version of the original paper. As it is, most people must continue to attempt to reconstruct the theory from the short and incomplete accounts of it which appear in different articles. Many of these are unfortunately so brief as to be almost incomprehensible.

The 33 articles are grouped into seven sections on general theory and basic experiments; measurement; expanded observations; physiological applications; recognition; frequency analysis; and speech communication. For each section, Swets has provided very short introductions, which make no attempt at evaluation and which collectively do not contain very much more than is already given in the table of contents.

Collections of papers of this kind are becoming very common. They are welcomed because they are undoubtedly useful and convenient. But in most cases, as in the present one, a well written book, by a single author, covering the same ground, would be very much easier to read and to evaluate. A collection of papers, no matter how well selected, must be both incomplete and repetitive. This collection suffers from both failings, but minimally; and while we must regret the non-appearance of the book which Swets might have written, this volume does provide an opportunity for us to take stock of this very important and rapidly developing area of psychological research.

Perhaps the most impressive result of these applications of signal detection theory is the use of a common language by people working on such diverse subjects as visual and auditory thresholds, speech communication, thermal sensitivity, auditory masking and physiological recordings from the frog's retina. This common language has not merely been superimposed on a set of unrelated subjects, but has enabled people whose interests are related to see the relationships more clearly.

However, others before and since Peterson, Birdsall and Fox have applied versions of statistical decision theory to perception. Almost no reference is made to their work in this volume. Notable examples among the neglected are Thurstone, Hecht and, more recently, Barlow. Such parochialism is a serious defect in the work of this group.

When one considers the details of the theory which distinguish it from other applications of statistical decision theory, the record is rather mixed. In the controversy concerning the nature of the sensory threshold, the evidence on the whole seems to favour a movable criterion rather than a high fixed threshold. But there are a great many loose ends which suggest that this is not the whole story. For example, the agreement between *d′* calculated from Yes/No and from forced-choice experiments is fairly good, but, when the relationship between *d′* and signal strength is determined, the two methods give different results (Swets, chapter 6 in this collection). And Barlow has shown that the efficiency of a Yes/No judgement is reduced when attempts are made to raise or lower the judgement criterion by the methods which are standardly used in the determination of *d′*.

A major difficulty for the theory is that while it predicts that *d′* should be proportional to the square root of the signal strength (E/No), most workers on hearing have found it to be linearly related to E/No, while in vision Tanner and Swets found *d′* to be proportional to the square or the cube of the signal intensity. This difficulty is not properly discussed in any of the papers in this collection, although Tanner (chapter 16) attributes it to the effect of uncertainty about the signal.

Unfortunately, the role played by uncertainty about the nature of the signal in determining the relationship between *d′* and signal strength is not properly discussed in any of the papers. Green (chapter 2) has a very brief discussion of the effect of signal uncertainty on the shape of the frequency of response curve. If his argument is correct, the conditions under which the subject has full knowledge of the nature of the stimulus can be achieved, and it is unfortunate that the relationship between *d′* and signal strength has not been determined under these conditions.

An important feature of the Peterson, Birdsall and Fox theory was the assumption that the subject tries to maximize his expected value. No serious attempt has been made to test this assumption. Green (chapter 2) has shown that when stimulus probability alters, the criterion does not shift as much as it should; and although one must assume that many R.O.C. curves have been obtained by altering a pay-off matrix, the relationship between the predicted effect of this variable and what in fact occurs has never been reported. It may well be that it is more sensible to treat this aspect of the theory in a non-quantitative way. But the arguments for and against this view do not appear in this volume.

One could go on in this way through many of the chapters of the book, bemoaning the scattered nature of the evidence presented, the failure to consider the results of other workers, and the lack of any clear exposition of some aspects of the theory. It all adds up to the conclusion that the time is ripe for a book summarizing and evaluating the work of this school. Until someone has the energy and courage to attempt such a book we must be grateful to John Swets for giving us this collection in a compact, though expensive, package. It makes tough and indigestible reading and most of it is suitable only for postgraduate students and other specialists. But it does represent very well the weaknesses and exciting strengths of the work of this school.

C. I. HOWARTH

Vision and Visual Perception. Edited by CLARENCE H. GRAHAM. New York and London: Wiley. 1965. Pp. vii + 637. 180*s*.

All six contributors to this volume are psychologists with long experience in vision research. Under the guidance of Clarence Graham, who has written nine of the book's twenty chapters, they have attempted to produce a work that is broad in scope and 'of a depth appropriate to advanced students'. *Vision and Visual Perception* aims to include 'considerable discussion of problems that mainly interest the psychologist', as well as discussing vision from the point of view of the physicist, physiologist, psychophysicist and psychophysiologist.

The subject-matter has been divided in a conventional way into the various forms of visual discrimination (brightness, flicker, colour, form, etc.), visual adaptation and contrast, with chapters on anatomy, electrophysiology, photochemistry and optics. Not surprisingly for a book that has been written by psychologists, *Vision and Visual Perception* is stronger on the behavioural than on the physical and physiological aspects of vision. The only chapter in the 'basic sciences' group that can be recommended as an authoritative review is the one on electrophysiology. Although much important work has been left out (on movement detectors, centrifugal activity, the early receptor potential, etc.), what has gone into this chapter has been well digested and clearly presented.

The best clue to the psychological content of the book is to be found in chapter 3 where the psychophysical methods are described as 'occupying a central role in visual experimentation'. These give data of two types: the 'stimulus-response' function, showing how some measured aspect of response varies as a function of some measured aspect of a stimulus, and the 'stimulus-critical value' function, showing how thresholds or critical stimulus values vary as functions of other stimulus variables. As a source-book of behavioural research that fits these formulae, *Vision and Visual Perception* is unusually extensive. On problems of perception that fall beyond the scope of these formulae (e.g. the constancies, visual appearances, illusions, ambiguous figures, figure and ground, and attention) this book has little to offer and nothing that seems 'of a depth appropriate to advanced students'. (A three-paragraph review of illusions, devoted mainly to a description of the Müller–Lyer and other well-known 'figures', hardly ranks as a study in depth!) Despite the authors' intention to cater for psychologists as well as psychophysicists, the psychology in this book is limited to psychophysics. And it is a psychophysics that will seem outmoded to many experimental psychologists, based as it is on an S-R model that takes no account of the new way of thinking about problems of visual discrimination that has come from communication engineering. In particular, it fails to consider the implications for the concept of 'threshold' of the idea that sensory discrimination is limited by noise in the nervous system. It is also a psychophysics confined almost entirely to man.

Vision and Visual Perception should prove very useful to the postgraduate student as a source-book of psychophysical (S-R) data and of the theories that have been proposed to explain them. But if he decides to purchase this expensive book he should be clear that he is buying what the editor aptly calls a 'sort of inventory'. If he expects something more exciting than this he will be disappointed. He will find that the authors seldom give any clear hints as to what their lists of facts might signify. It is the reader who must discover that the experiment on the images of metameric lights, filed systematically under 'color characteristics' on page 490, has implications for the three-channel hypothesis of colour vision described on page 414, and that the experiment by Rushton and Westheimer referred to on page 204 has implications for Wald's 'compartment hypothesis' of dark-adaptation described on page 193. And it is the reader who must decide, after ploughing through masses of experimental facts and detailed descriptions of theories, which, if any, of these theories can be excluded. Because of the almost complete absence of critical discussion and evaluation and the persistent shelving of all theoretical and controversial issues the book makes dull reading and seems unlikely to arouse much enthusiasm for the subject in students. It is the kind of useful book to which they should be exposed only when they are very 'advanced'.

G. W. GRANGER

Probability and Statistics in Psychological Research and Theory. By D. W. STILSON. San Francisco: Holden–Day. 1966. Pp. xii + 507. $12·65.

This book is a well-constructed, insightful, and at times exciting introduction to the use of Bayesian statistics in psychological decision-making. In addition, it will serve as a useful textbook of elementary probability theory for psychologists in all branches of the subject.

All introductory texts on probability theory must cover much the same ground, with variations only in style; the treatment here is pitched neatly between algebraic sterility and tedious circumlocution, and will appeal for this reason. The author states a preference for intuitive clarification over algebraic argument, and in the former respect deals excellently with the nature of statistical thinking, the meaning of probability, the link between Bayesian and psychological thought, etc. The algebra is concise and not over-demanding; inessential proofs are relegated to appendices, or references given to where they may be found.

Nevertheless, the book does not quite live up to the expectation of its title. It will be read and used by first-year undergraduates and research students who will already have encountered the commoner statistical methods. The reader is taken to the brink of understanding the theoretical basis for these, but that is all. For example, the derivation of confidence-limits for means is stopped just short of the use of Student's t distribution, discussion of the chi-squared distribution in relation to testing hypotheses about variances does not mention the F-test for variance ratios, and applications of exact binomial probabilities are given without mentioning the common non-parametric tests based on them. These points should not escape an intelligent student, but an opportunity of linking the well-argued theory to the often obscure practice seems to have been lost.

The examples and problems in the text are interesting and instructive. Suggested readings are given mostly in the form of page or chapter references, which is a great help, although it is a pity that the author-names of these are not given in the main index.

D. H. TAYLOR

Mathematical Explorations in Behavioural Science. Edited by F. MASSARIK and P. RATOOSH. Homewood, Illinois: Rd. D. Irwin and Dorsey Press. 1965. Pp. viii + 387. $7.95.

The collection of papers by various authors which comprise this volume covers a fairly wide range of topics, from individual behaviour in the detection of repetition to mass behaviour in attendance at national parks. The connexions of the book, however, as may be seen from the titles of the other volumes in this series, are with management and organizational science, and the greater part deals with group performance. The degree of mathematical treatment varies very considerably and is non-existent in some of the contributions—e.g. R. L. Maier on information overload in institutions. There is little which is completely new in the book but it does provide interesting reading for anyone attracted to a rigorous approach to group and mass behaviour. It offers very little to anyone more concerned with individual processes.

There are some introductory and methodological papers, of which Hunt's discussion of an amplification of Bayes' theorem is the most important, since it deals with the perennial and perplexing problem of model evaluation. The short paper by Harrah, on the logic of semantic information, will be very difficult to comprehend by anyone except the specialist and appears to demand a knowledge of his book: it should therefore hardly have been included here. A similar criticism might be applied to that by Guetzkow, under group behaviour, which, although a very relevant topic (Computer-aided simulation of inter-nation decision-making), 'assumes some familiarity' with his other publications and this turns out to be a considerable under-assumption.

The part on 'individual behaviour' contains thirty non-mathematical pages, of which twenty are taken up with a paper by Miner on conformity to national and organizational norms. Hence, Attneave's contribution to this part (Detection of repetition) is brief and Colby's is even briefer.

The remainder of the volume is generally substantive and interesting. This contains the parts on group and large social aggregate behaviour. An experimental study of two- and three-person games by Lieberman indicates some of the insufficiencies of the relevant theory and H. Soloman also uses experimental data in his discussion of the model of Lorge and Soloman of the effect of group size on verbal learning. Coleman attempts to describe diffusion in incomplete social structures and, incidentally, gives a good indication of the severe restrictions imposed by the mathematical treatment of events in this field. Bartos has a model for a particular negotiation situation and experimental data help him to identify some of the problems involved.

Attendance at national parks is dealt with by Catton in terms of gravitation models which clearly exemplify the 'social physics' approach. It may well be that in the behaviour of large social aggregates certain simple laws of this kind will apply and it is up to the mathematical social psychologist to discover them. Although one must not expect that the laws of physics will apply in their simple form to the parameters of a society, there is clearly much to be gained by a comparison of societies with the behaviour of large collections of particles of one kind or another.

The book has its disappointing features, particularly for those primarily interested in individual processes, but it should be useful to anyone new to the use of mathematics in social or management science.

A. R. PIKE

Readings in Animal Behaviour. By THOMAS E. MCGILL. New York: Holt, Rinehart and Winston. 1965. Pp. x + 592. $12.00.

The subject of animal behaviour is part of several disciplines. This book of readings brings together reviews and summary accounts of research programmes, as well as reports of specific experiments in each of these disciplines. Fifty-five papers have been selected from a wide variety of journals, ranging from *Science* to the *American Journal of Orthopsychiatry*, and unlikely to be all available in most universities or research institutes.

The articles are grouped under seven headings: introductory readings (5); behaviour genetics (6); neural, hormonal and chemical control of behaviour (8); the development of behaviour, critical periods and imprinting (11); sensory processes, communication and orientation (8); learning and motivation (10); social behaviour, ethology and evolution (7). Each section is preceded by an editorial introduction in which the problem is set out, terms used in the particular discipline are explained and general references given for further reading. In the introduction to individual papers, there is again emphasis on definition of terms. References are given to other major articles by the author and the current interest of the research is outlined. The volume also has an author and a subject index. A major feature is that thirty-eight of the articles have been published since 1960 and only three before 1955. The choice of papers in each area has necessarily been selective but generally covers the major contributions in the field. Occasionally, these are still the subject of heated controversies, as in the case of the transfer experiments involving planarians. Here, a reference to a more general review of the problem might have been valuable.

The broad coverage of recent studies in animal behaviour and the high proportion of review articles in this book make it invaluable as a compact source of information.

DAPHNÉ JOYCE

Current Research in Motivation. R. N. HABER. New York: Holt, Rinehart and Winston. 1966. Pp. xii + 800. 100*s*.

There may be some initial disappointment at finding that this is not an original book but a book of readings. Of course, in principle, we welcome books of readings, but feel slightly cheated if the fact is not apparent from the title. However, once over this disappointment, and having started out to glean a preliminary impression of the tome, the reader may well find himself engrossed. It says much for the editor that his chosen authors, covering a wide variety of specialized interests, give on the whole a lucid account of motivation theory and research.

Certainly the aim of presenting thought and experiment on motivation—interpreted widely—has been achieved with some success. There are 76 papers here, including some not previously published or obtainable only with difficulty. They are organized into nine chapters overlapping somewhat in subject-matter but covering the topics of instinct, drives, punishment, arousal, exploration, incentives, conflict, anxiety, and fantasy in relation to motivation. Each chapter is preceded by an explanatory note by the editor, who has not always succeeded in his intention of integrating the various works chosen, and whose introduction is often little more than a catalogue of the ensuing chapter.

The strength of this collection lies in editorial judgement in choosing material. We are presented with an absorbing and fairly complete picture of the field of motivation almost up to the minute, with writings from 1949 but mainly in the period 1958–65. To lengthen the period under review, Elizabeth Duffy, D. O. Hebb and J. B. Gilmore provide mature and historical perspectives on work extending back over the years. It is saddening to find British psychology as poorly represented (only an excerpt from J. Bowlby's monograph, *Maternal Care and Mental Health*) but this is a reflexion on British psychology rather than on the editor, though *Experiments in Motivation* (Eysenck, 1964) must have appeared in time for consideration. Haber has not been concerned to emphasize one school over another, nor has he been afraid to draw attention to controversy (papers and letters by W. A. Hillix and by G. P. Sackett).

Contributors are so numerous that only those that gave particular pleasure can be mentioned. N. E. Miller takes us with enthusiasm and at great speed through a series of ingenious and exciting experiments on alimentary sensory feedback (providing evidence on determinants of the effective point of reinforcement in a temporal sequence), and another well-known section on conflict behaviour and drugs.

J. McV. Hunt is thorough as usual on the epigenesis of intrinsic motivation. There are various discussions of the inadequacies of the multiplicative drive concept (and hence the emergence of the recent interest in exploratory behaviour) as well as of drive-reduction theory, adaptation-level theory and drive-induction theory.

While there is a nice apportionment of theoretical and experimental reports throughout, the later chapters tend to show the increasing interest and experimental sophistication of work on human motivation, while the early chapters are predominantly based on animal experimentation, from insects through the never-to-be-forgotten rodents to monkeys. Here we find the meticulous attention to experimental detail on which the comparative psychologists have built their reputations.

Excited by much of what had been offered, the reader may reach chapter 9 (on 'Fantasy, dreams and unconscious processes') hoping for further enlightenment on topics of clinical relevance. But this is in fact the most disappointing chapter. It is here that theory is least clear and least profound. It is here that experiment is easiest to criticize. The inclusion of R. A. Clark's inadequately controlled experiment on the effects of sexual motivation on fantasy indicates the paucity of work on a sound methodological basis in this area. At a time when clinical psychology is returning to the animal laboratory for a refresher course in experimental methodology and for increased precision in solutions to the problems presented by disordered human behaviour, we should expect only the best as exemplars in a collection of works such as this.

While the body of the book gives a satisfying feeling of activity to some purpose, one is still left with the impression that little can be offered in the way of clinical relevance. We still lack statements of value to the clinician. What can we say has gone wrong in the motivational history of the middle-aged man who announces, after years of steady bread-winning, that he prefers tramping to working for his family? What can be suggested in the management of an intelligent young girl who responds well and normally to social stimulation and other external motivation, but, left to herself, remains motionless, even neglecting her own feeding and hygiene? Why is it so difficult to arouse general motivation in an adult outside a Skinner box or other laboratory setting?

D. C. McClelland's interesting section on the importance of early learning, though directed towards explaining the persistence of motivations once aroused, goes some way towards answering the last question.

Haber offers a book which will be invaluable to senior undergraduates, researchers and teachers, but Holt, Rinehart and Winston have produced the book with some evidence of haste. Misprints are noticeable; references are excluded from individual articles and accumulated at the end; there is no author index and the general index is meagre for such a volume.

ANNE BROADHURST

Evolution and Growth of Human Behaviour. 2nd edition. By N. L. MUNN. Boston: Houghton Mifflin. 1965. Pp. xi + 594. $8.50.

Child Psychology: Behaviour and Development. By R. C. JOHNSON and G. R. MEDINNUS. London: Wiley. 1965. Pp. x + 556. 53*s*.

These books present two very different approaches to the study of behaviour development and child psychology. The ten years which have passed since the revised edition of Munn's book appeared in 1955 have seen very considerable advances in a number of fields which bear on the central theme of work; in particular the recent research on the genetic basis of behaviour and the ideas this has generated about the evolution of behaviour. Each of the chapters has been revised, though in some cases, as the author points out, so little new work has been done that a revision is barely necessary. The chapters on unlearned behaviour, prenatal behaviour and the development of motor skill provide examples of these. The bibliography has been extensively revised and, though not exhaustive, is very adequate. In the course of the past decade this volume has fulfilled a need effectively; this revision should enable it to continue doing so, though not, one hopes, for a decade.

Johnson and Medinnus divide their book into five sections: introduction (1 chapter); basic factors in development (5); the family and its influence on development (4); societal influences on socialization (3); the end product (4). The introduction provides some discussion of the methods and techniques of child psychology. Here is a very marked imbalance: for example, two pages are given to a quotation from G. S. Hall, whilst concepts such as validity and reliability get only two or three short paragraphs. On pages 7 and 8 there is a most misleading definition of genotype, which is followed by an equally facile discussion on the merits of the genotypic versus the phenotypic approach to the study of behaviour.

The basic theme running through the book is a polarity between 'biological-genetic' forces, which the authors see as responsible for similarities in behaviour, and social forces, which are seen as being largely responsible for producing differences in behaviour. The second section contains a chapter on heredity which is adequate, though no emphasis is placed on the developmental effects of genetic variables. The chapters on 'growth and maturation', 'learning and motivation', 'language' and 'intelligence' do not provide adequate coverage for an undergraduate course. In this section on basic processes no mention is made of perceptual development, motor development or the development of spatially co-ordinated behaviour.

The sections on the influence of the family and 'societal' influences on the socialization of the child are good. In each case an up-to-date and exhaustive treatment of recent work is provided. The final section contains a rather pedestrian chapter on behaviour disturbance, including two pages on how to avoid quacks. The chapter on personality development is disappointing.

The book cannot be recommended as a text on child psychology for undergraduate or for education students; far more satisfactory would have been the publication, as a separate monograph, of the sections on social and cultural factors.

KEVIN CONNOLLY

The Psychology of Classroom Learning. By JOHN M. STEPHENS. New York: Holt, Rinehart and Winston. 1965. Pp. 511. 60*s*.

For an increasing number of students in colleges and departments of education J. M. Stephens' earlier work—*Educational Psychology: the Study of Educational Growth*—has in recent years served as a major text-book. The word gets around that the content is comprehensive enough to cover the more important topics in the course and that the author's urbane style makes his work eminently readable. What is also keenly appreciated by those students who have discovered Stephens after examining some of the other texts that are available in this field is that he appreciates and goes a long way towards satisfying the needs of intending teachers. The majority of such students are interested in theoretical formulations and the data supporting them only in so far as these involve recognizable implications that can serve as principles to guide classroom practices. Stephens clearly sympathizes with this expectation and sets out to gratify it, but without jeopardizing his integrity as a scholar in the process. He is willing to allow what he describes as 'fireside overtones' to intrude into his discussion of the implications of research data, but always remains close to the available evidence and, where necessary, exposes its inadequacy or inconsistency.

This more recent work contains, intact, some of the sections from the earlier volume, but to describe it as a revision would be seriously to underestimate the amount of new material that has been included and the extent to which most parts of the book have been remodelled. There are three main sections. The first part deals with the role and functions of the teacher, includes a compressed account of child development—with ample references designed to indicate that this important topic should be followed up elsewhere—and a chapter on educational measurement which also needs to be considerably supplemented from the sources. Part 2 is the section to which the attention of students could be most profitably directed. The nature and conditions of learning are here presented with commendable clarity and economy and a good deal of recent (post-1960) data is included. Part 3 is concerned largely with character formation and socialization. A short, final section deals with some of the administrative devices—grouping procedures, team teaching, etc.—which affect the relationship between teacher and taught and ends with a chapter in which are outlined the frustrations that teachers regularly suffer and the satisfactions they occasionally enjoy. There is also an appendix entitled 'Some elementary statistics' which is too brief to be of value.

A feature worthy of mention is the inclusion at regular intervals of insets containing accounts of original investigations that are relevant to the theme of the surrounding text. These are presented without comment or interpretation and may therefore be used, not only to supplement the exposition, but also, as the author suggests, 'to remind the student that the interpretation of research is difficult and precarious and that definite advice does not emerge automatically and unambiguously from the data themselves'.

The references are ample and somewhat more eclectic than those supplied in many American texts: a sizable number of articles from this journal are quoted.

ALFRED YATES

Reading Disorders: a Multidisciplinary Symposium. Edited by R. M. FLOWER, H. F. GOFMAN and L. I. LAWSON. Philadelphia: Davis. 1966. Pp. x + 146. 24*s*.

This little book is a collection of papers from a symposium held at San Francisco Medical Center; contributors included professors of education, psychologists, professors of paediatrics, psychiatry, neurology, opthalmology, audiology and speech disorders. The contributions are of a very uneven quality. The book opens with a wise and succinct paper by a paediatrician, which is a good summary of present thinking on the subject. She makes the point that a psychologist is an essential member of any team dealing with children having reading disorders. Unfortunately the next paper, from a psychologist, falls well below the standards of the best practice in educational psychology. The psychologist seems limited to a repertoire of three or four tests and to be unaware of the variety of useful tools available for use in diagnosis and assessment. Recent tests such as the *Illinois Test of Psycholinguistic Abilities* are ignored. Too much weight is given to W.I.S.C. profiles, and the common fallacy that the sub-test differences indicate differences in specific factors seems to be implied. The reviewer submitted the results quoted to a Maxwell Factor Analysis and found the difference between verbal–intellectual and space-performance factors to be completely insignificant. Perhaps the best contribution on the psychological side is that of Marianne Frostig on teaching reading to children with perceptual disturbances.

As far as a psychologist can judge, the opthalmologist and audiologist have produced well-balanced and useful papers. However, the article on remedial teaching adds nothing to what the most newly qualified teacher should already know.

Many papers in this symposium adopt the views of the 'word blindness' school with respect to reading difficulty. The usual references are given in support of these views, but none in favour of alternative hypotheses. Again, the well-known investigations, such as those of Hallgren, supporting a genetic basis for reading difficulty are quoted, but there is no mention of the great weakness in design of many of these investigations. It is not revealed that theories have been constructed on the basis of concordance/discordance of just three pairs of identical and three of fraternal twins. One has the impression that clinical hunches of a child's intelligence or reading ability have been accepted frequently without any objective measurement.

Read critically, this book could be a useful source of information, but the standard of the work is not that normally expected from a professional symposium.

ALEC WALBRIDGE

Behavior, Aging and the Nervous System. Edited by A. T. WELFORD and JAMES E. BIRREN. Springfield, Illinois: Thomas. 1965. Pp. 637. $22.50.

Human ageing as an area for experimental psychological research is well provided with periodic reviews of advances in research and theory—Welford's *Ageing and Human Skill* and Birren's *Handbook of Aging and the Individual* being two notable examples. These two authors have now collaborated to edit the proceedings of a colloquium on age changes in speed of behaviour. This theme is the focus of the papers presented here, although the topic is treated from several different points of view. Thus there are contributions from physiologists, neurologists and clinicians, as well as experimental psychologists.

The general plan of the book is that papers on age changes in psychological functioning are presented first; these are followed by reports of physiological studies whose outcome could shed light on the mechanisms involved in the slowing of behaviour; finally several papers are presented which describe behavioural slowing in conditions other than ageing. In the first section experimental work is described on age differences in decision processes, vigilance, perceptual discrimination and short-term memory as well as studies more directly concerned with age-related slowing in speed of performance. The physiological studies reported are concerned with such issues as age changes in the autonomic nervous system and in level of arousal, relationships between E.E.G. measures and performance, age changes in neural function, and generally with the causes and effects of slowing at the physiological level. The third section includes papers on slowing due to cardiovascular pathology, brain damage and schizophrenia—there are some interesting parallels to the effects of ageing here, although several of the authors warn against over-simple conclusions being drawn with regard to common mechanisms being responsible.

This book should prove a rich mine to be worked by those interested in problems of ageing, cerebral dysfunction or speed of behaviour. It is a most useful collection of data and techniques and there has obviously been a fair amount of cross-fertilization of ideas between workers in different disciplines.

The valuable thing about this collection of papers is that it provides a coherent body of experimentally obtained facts on age changes in brain function and behaviour. Paradoxically, one of the book's least satisfactory features is the very empirical nature of almost all the studies reported, which deprives the book as a whole of an integrated theoretical framework. Despite the editors' decision not to include a discussion of trends in research, it might have been useful from the reader's point of view if more space were taken up by articles putting the various lines of research into some framework and attempting to interrelate the major findings.

Nevertheless, it should prove a valuable source of reference and stimulation to experimentalists working in the fields of ageing or C.N.S. dysfunction.

F. I. M. CRAIK

The Social Psychology of Social Movements. By HANS TOCH. Indianapolis: Bobbs-Merrill. 1965. Pp. xiv+257. $6.50.

This book is for the undergraduate social scientist. It will hardly satisfy psychologists, sociologists, anthropologists or historians who have been professionally concerned with studying social movements or collective behaviour. It will mislead others if, in reading this book, they take it as representative of the depth of thinking in this field.

Its virtues as a teaching text are considerable. Many as well as Hans Toch took heart from Hadley Cantril's *Psychology of Social Movements.* Little has been published on this subject since then that has been capable of drawing a student from a concern with his society's problems to a professional concern with the problems of man. Toch does very well in this respect. From a wide and topical range of examples he proceeds without effort to involve the reader in scientific analysis, using a low-powered but very systematic conceptual framework. At each point where one feels that an obvious complication is being ignored, it turns up and is thoroughly dealt with. The overall framework is better than that used in most published studies of social movements but the book still warrants the restriction made in the first paragraph of this review.

Toch has made it easier for the beginner by attributing a rational problem-solving character to social movements which Everyman intuitively feels contain essential elements of madness and conspiracy. In so far as he agrees to the existence of these latter phenomena, Toch lumps them together (pp. 55–6) and deposits them in the extreme category of individual paranoidal behaviours. This is a helpful procedure during the first stage of professionalization, but constitutes a serious barrier at the latter stages where one can hardly avoid the feeling that Everyman was closer to the truth, even with allowances for ignorance, stupidity and self-interest. Toch's own examples give ample room for an alert lecturer to raise such questions as to whether the emergence of the leadership and their following are not independent processes that interact by means of appeals, etc., whether the madness—the role of unconscious psychological factors—is not so much a part of the led as of the leaders, of the fellow travellers as of true believers. Toch might of course be right in his implicit assumption that these questions cannot be handled with students until there is some faith in the possibility of a scientific solution.

F. E. EMERY

Problems of Television Research: A Progress Report. Television Research Committee. Leicester University Press. 1966. Pp. 38. 3*s*. 6*d*.

Psychologists and other social scientists rightly accorded a warm welcome to the first of the Television Research Committee's Working Papers when it appeared in 1964. Entitled *The Effects of Mass Communication, With Special Reference to Television*, it turned out to be a first-class summarization by J. D. Halloran of the research results achieved in the field of mass communication and of the theories developed from these results. For those marginally involved in working with the mass media, e.g. in agricultural advisory services, good-cause promoting, etc., it could not be too strongly recommended, though it could never, of course, replace the more substantial works as standard texts for the serious research worker.

It is less easy to accord the same warm welcome to this second production of the committee, not because it is unhelpful or misleading—it is certainly neither—but because to the experienced research worker the fact that it needed to be written at all is depressing. Though this booklet does not bear the name of Halloran as author, it is presumably largely his work, designed to continue the education of the committee of which he is secretary. This 'progress report' gives a comprehensive account of the types of research strategy that its author believes—as do many more of us—will produce useful contributions to the understanding of how mass media and society interact. So far so good. But what we really need, the researcher will say, is less talking and more doing. It is more than twenty years since Berelson, Lazarsfeld, Merton and others did their thinking about research strategies, almost as many since Hovland and his colleagues at Yale devised theirs, ten since Festinger, Osgood, etc. studied attitude formation and change anew. Surely it is time we were less apologetic about the limitations of the tools of mass communication research, and certainly high time we gave up saying how difficult it all is.

Good luck to Mr Halloran in his valuable task of teaching the Home Office Committee the bases of social research but do not let anyone imagine from this book that, because of limitations in research techniques, no useful work can be done. Much has already been achieved, outside the auspices of the committee, and more will continue to be done all over the world by an increasing number of researchers in many disciplines. The book specifically mentions the need for an Institute for Mass Communication Research. By all means—but meanwhile the money so generously provided by the ITA is stagnating, whilst the committee, Micawber-like, wait for something special in the way of research plans to turn up.

B. P. EMMETT

Human Robots in Myth and Science. By JOHN COHEN. London: Allen and Unwin. 1966. Pp. 156. 35*s*.

Dr Cohen has here compiled a most comprehensive history of robotry, ranging from the oracular busts of ancient Egypt, through the Ars Magna of Ramon Lull, to Vaucanson's duck (a fascinating beast complete with simulated digestive system and an anatomically correct wing structure) and Babbage's analytical engine. Since a detailed account of robots over such a time span would require many volumes, the author has chosen 'to gaze at [automata] from a respectful distance' rather than present the microscopist's view of the subject. However, the lack of detail inherent in this broad approach may irritate the casual reader.

The book comprises nine chapters, the first seven of which are devoted to historical development, culminating in a discussion of the logical and mathematical machines of the late-eighteenth and nineteenth centuries. Dr Cohen presents a bibliography and reference list at the end of each chapter.

Considering the impact of machines on social life, the author briefly discusses the resistance to new machines and ideas displayed by ordinary folk, particularly in Britain, and quotes Stendhal (1826) that 'England's basic moral theme is the abhorrence of machinery'. Unfortunately he leaves unanswered the question of a psychological palliative for those who feel offended by new ideas or threatened by new machines.

Dr Cohen has explained a little of the relationship of man to the machines he makes in his own likeness. He has sought to trace the history of the idea that a synthetic man, call it automaton or robot, can be made. In this idea he finds a link between the animate and the inanimate, the organic and the inorganic. The desire of the artist, not merely to portray man but to create man, is manifested in the urge to build robots.

Dr Cohen has understressed the magnitude of the leap from the idea that man-like statues could be magically imbued with mind to the idea that a machine does not necessarly need to look human in order to have mind-like properties. The logic machines and calculators of Lull, Pascal, Leibniz, Pierce and Babbage are to the cybernetic revolution what the wheel was to the industrial revolution.

Dr Cohen has approached the subject of robots on a very broad front and produced a most stimulating book. I hope he finds time to fill in the details in later works. G. L. MALLEN

Psychopathology: Its Causes and Symptoms. By F. KRÄUPL TAYLOR. London: Butterworths. 1966. Pp. ix + 356. 70*s*.

Professor G. M. Carstairs tells us in a foreword that this book gives us an exceptionally lucid account of the subject-matter of psychiatry and the present limits of knowledge in the speciality. Yet your reviewer has found it a curiously difficult book from which to extract information or explanations or, despite the 545 references, to get leads into the recent literature.

Dr Kräupl Taylor has been at great pains to clear the ground thoroughly first. Thus in Part 1 he re-examines at some length the concepts and terms which have been used in medical writings about clinical symptoms and diseases, and discusses the relationship of psyche and brain, the differences between objective physiological processes and subjective psychological phenomena, the difficulties of nosological classifications, and the genetic, physiogenic and psychogenic aetiologies of disease.

Part 2 is taken up with descriptions of disorders of perception, intellect, affect, mobility, memory and consciousness. Agnosia, dyslexia, illusions and hallucinations, for instance, are described; on the other hand, little is said about the conditions under which they occur. The inclusion of a classification of sexual perversions gives this part a nineteenth-century flavour. Mention is also made of the deficiences of the mnestic functions, such as cryptomnesia and paramnesia. Part 3 expounds the essentials of modern dynamic psychopathologies, and discusses trance states and hysteria, the development of Freud's theories, and his influence on the psychological thought of today.

The term 'psychopathology' seems to be coming back into use, and during the last few years several books on psychopathology have appeared which apply the knowledge gained from experimentation in psychology to the explanation of disorders of behaviour. Whatever its shortcomings, the modern experimental approach, which makes it relatively easy to define terms operationally, would have saved Dr Kräupl Taylor from what must seem to the experimentalist to be rather protracted manœuvrings for position before launching an attack on the practical problems. His book has grown out of a different and older tradition. It is not without interest for all that. D. RUSSELL DAVIS

An Experimental Approach to Projective Techniques. By JOSEPH ZUBIN, LEONARD D. ERON and FLORENCE SCHUMER. New York and London: Wiley. 1965. Pp. xx + 645. 105*s*.

This work comprises twelve chapters and an epilogue. Most of the chapters end with a convenient summary of findings, arguments and conclusions.

In the first chapter the authors note the challenge of projective techniques as something outside the mainstream of scientific psychology but yet deserving evaluation in the most objective way possible. The next two chapters examine the general position that perception is an approach to personality; the authors summarize thirteen major theories of perception and five major approaches to personality through perception. Their conclusion is that a direct relationship between perception and personality is difficult to demonstrate; we all live in different perceptual worlds but a relationship between this fact and personality differences has not been established. In applying their conclusion to projective techniques the authors affirm the view—expressed by Zubin elsewhere—that such techniques are not perceptual tests in the traditional sense.

Four subsequent chapters are devoted to the Rorschach technique and three to the Thematic Apperception Test; one chapter concerns three derivatives of the Rorschach and the remaining chapter deals with ten derivatives of the T.A.T. Both the Rorschach and T.A.T. are discussed in detail under the same headings of 'Historical perspectives', 'Current status' and 'Methodological problems'; each is evaluated at length as a psychological experiment and a careful psychometric approach is applied to each. In the case of the Rorschach the psychometric approach culminates in the presentation of an elaborate rating system by which the responses are assessed under 10 principal categories and 85 subordinate categories. The psychometric approach to the T.A.T. produces two general rating scales for emotional tone and outcome, specific rating scales for emotional tone in the response to each of the cards, a check-list of 163 themes classified under two principal and various minor categories and other check-lists for unusual formal characteristics and usual deviations in response to each card.

In the epilogue the authors conclude that any attempt to eliminate projective techniques would only provide the need to smuggle them in through the back door. Asking whether the usefulness of such techniques has been established, the authors answer: For some limited purposes, perhaps, Yes, but for the general purpose of evaluating personality the answer must be very tentative and probably negative. Projective techniques were once young and promising, they say: today they are only promising. Clinicians will give up their interest in these techniques only if provided with better tools. They look forward to the time when some of the claims and clinical hunches which experimentalists now disdain become the cornerstones of new objective findings made available by the more powerful tools of the future.

This work is unlikely to have much appeal to practitioners. Its warnings will not influence those who employ projective techniques, nor will its suggestion that these techniques might be developed into objective, psychometric instruments attract to their use those who at present reject them. For those interested in certain theoretical problems—e.g. inkblot perception as a special form of pictorial perception and the production of narratives in response to pictures—the work is a very rich source of material. One would particularly like to commend here the analysis of the two major techniques as psychological experiments and the meticulous psychometric approach applied to each. In relation to inkblot perception Zubin appears to have retreated in some measure from his 1954 position that the Rorschach is essentially an interview conducted through a veil of inkblots and to allow that there are important stimulus factors to be investigated along Gibsonian lines. Independently of these rather specialized problems, the succinct summaries of perceptual theories and perceptual approaches to personality are of intrinsic value.

JAMES PATRICK S. ROBERTSON

Clinical and Social Judgment. By J. Bieri, A. L. Atkins, S. Briar, R. L. Leaman, H. Miller and T. Tripodi. New York: Wiley. 1966. Pp. 271. 60*s*.

The purpose of this book deserves a round of applause. It tries to utilize ideas and findings from the experimental psychology of judgement to cast light on how clinical and social judgements are made; in so far as psychologists make clinical and social judgements it is reflexive in intent. A further merit is the thoroughness with which the literature has been searched and summarized. It is a workmanlike and well constructed failure.

Part of the failure is due to two forms of theoretical indiscipline. The authors collect concepts from a variety of frameworks—psychophysics, information theory, construct theory, decision theory (cognitive dissonance seems to be the only missing candidate)—but neglect to say how or whether these diverse notions are to be welded into a coherent viewpoint. The 'integration' offered is of the scholarly-gossip type. The criterion for integration is as for putting knick-knacks on the mantelpiece—they look well together. Sometimes cafeteria selection from disparate theories is abandoned in favour of *ad hoc* empiricism—e.g. 'the lack of attention given situational factors in judgement theory and research makes it desirable that these factors be considered separately at this stage in the study of judgement, regardless of where the principle of parsimony ultimately will locate these variables'. This reflects the pious hope that ultimate theory will disentangle earlier conceptual confusions and the habit of trying to excuse an ill-formed emergency concept by calling it a 'variable'.

Aside from repeated mixing of theoretical metaphors, the pay-off from each separate line of attack is meagre. The extensive discussion of information theory models is followed by reports of studies in which subjects categorize fabricated case-histories of various types along prescribed dimensions (degree of defensiveness, clinical diagnosis, etc.) on the lines of psychophysical experiments.

Whether or not inference from the fixed-array, fixed-dimension model to clinical judgement is ultimately meaningful or not, the authors clearly fail to make anything of it. They conclude, for example, that the capacity of a judge to discriminate among arrays of social or clinical stimuli is not large. This turns out to mean that if we ignore unmeasured and unexplained individual variance, limit our judges to a response system which we invent and they must work to, make this a simple response system, supply a fabricated rather than a situation-generated series of social stimuli, subtract a constant term which we admit may be arbitrarily homogenizing transmitted information scores, then we may conclude that...

From personal construct theory Bieri draws the notion of cognitive complexity–simplicity as a dimension, the notion that subjects may be making their judgements in terms of either a few clustered dimensions or in terms of a complex series of relatively orthogonal dimensions. His own work, as reported, seems to have moved steadily away from construct theory in that he talks of cognitive complexity–simplicity as a trait rather than as a varying process with an individual; this in spite of the work of Runkle and Damrin showing predictable intra-individual movement along this dimension over time. Additionally, he ignores the construct theory implication that the complexity–simplicity level of different sub-systems within the same personal construct system may differ as an aspect of varying ranges of convenience.

The language of the book is in the tradition of psychological texts—doughy. Nothing is simple if it can be simplistic.

On reflexion, the poverty of implication from the psychology of judgement to clinical and social practice reflects only partly the limitations of the authors—it reflects much more the poverty of psychological work on thinking and judgement. Maybe the authors can claim to have done the best that might be done in the light of Hebb's still true dictum of 1949 that 'the failure of psychology to handle thought adequately...has been the essential weakness of modern psychological theory'. If so, categorizing clinicians who wish to amend their sinful ways would do better to by-pass psychologists for the time being and refer to the semanticists and logicians.

D. BANNISTER

Insight vs. Desensitization in Psychotherapy—An Experiment in Anxiety Reduction. By GORDON L. PAUL. London: Oxford University Press. 1966. Pp. 148. 40*s*.

Few controversies in psychology have aroused the degree of interest evoked by the dispute between the protagonists of the 'dynamic' and learning-theory approaches to psychotherapy. Strong feelings and personal commitments often obscure the very real academic and practical issues involved.

Paul's book is a distinct contribution in that he avoids polemics and special pleading and attempts to provide evidence respecting the effectiveness of the two therapies through direct experiment. His basic purpose is to compare what he calls 'insight' therapy with the behavioural technique known as desensitization, and to provide a reasonably objective basis for evaluating the results of the treatments.

The project involved seventy-four people, made anxious by speech situations, who had enrolled for a course in public speaking These subjects were allocated to one or other of five treatment or control groups on the basis of overt behavioural signs of anxiety and psychological test results. The measures were repeated following five 'treatment' sessions to evaluate changes attributable to these intervening processes.

Five experienced 'insight' therapists were employed to administer all treatments, this choice being made partly to offset any bias which might favour behaviour therapy.

The outcome was very clear-cut, Paul's summary of findings showing that, while 100 % of the desensitization group could be classified as 'improved' or 'much improved', only 47 % of the 'insight' treatment group could be similarly rated.

There are two fairly serious limitations of the experiment, and therefore of the book as a whole. First, it is difficult to see in what sense the experiment is a real test of 'insight' therapy, and many people will regard Paul's apparent use of the term as having little meaning or significance.

Secondly, none of the subjects was psychiatrically ill, at least in a formal sense, and the experiment does nothing to resolve the thorny problem of how the rival therapies compare for clinical patients. It may be that all Paul has demonstrated is that habit-training is most efficient where modification of habits is the problem.

Finally, while the material is well written and the findings reported are useful, one might reasonably question whether a book is the most suitable form of publication for what is very largely an account of a small-scale experiment.

H. B. BEECH

Psychology: The Science of Behaviour. By R. L. ISAACSON, M. L. HUTT and M. L. BLUM. New York: Harper and Row. 1965. Pp. ix + 374. 28*s*.

This is a somewhat different, but generally stimulating, text-book by three psychologists of rather disparate backgrounds. Isaacson has interests in both physiological and educational psychology; Hutt is a clinical psychologist and psychotherapist; Blum is concerned with industrial psychology and market research. Of some importance is the fact that the authors have explicitly restricted the scope of their book, leaving for another volume (in preparation) what they term 'the social-science aspects of psychology'.

Following an introductory chapter, the book pursues a course from the 'molecular' to the 'molar'. Chapter 2 deals with genetics at some length, appearing to be reasonably *au fait* with recent advances in our knowledge of RNA and DNA. In chapter 3 ('Biological foundations of behaviour') several important approaches often ignored by elementary text-books are included, such as McConnell's work on planaria, the bee work of von Frisch, and Olds' work on electrical brain stimulation. No extensive coverage of any of these approaches is given, however; while some doubt in cast on the view that Olds has hit upon 'pleasure centres', no alternative viewpoint is offered.

In chapters 4 and 5, perception is dealt with at length. The usual topics are covered, and, moreover, some of Granit's work is discussed, as is the Tanner–Swets (1954) decision analysis model. An exceptionally clear account of the latter theory is given. This account, together with a good analysis of Hebb's (1949) theory, means that the chapters on perception are both valuable and informative.

Chapter 6 ('Basic theories of learning') deals with most of the noted learning theorists of the 1930s. The chapter has several weaknesses: several pages are devoted to the minutiae of Hullian theory but none of the cogent criticisms of the theory (e.g. Koch, 1954) is included; no recognition is made of the fact that Tolman (1959) moved towards a 'reinforcement' theory; and the account of Guthrie's theory omits reference to the changes introduced after about 1940. On the other hand, mathematically formulated learning theories are dealt with well at a simple level, and mention is made of Grey Walter's models from neurophysiology.

Chapter 7 ('Verbal learning and higher cognitive functions') deals largely with memory in a very simple and somewhat inadequate way. Chapter 8 ('Motivation: views and theories') is fairly exhaustive. The views on motivation of Spence, Maier and Lewin are all given, but uncritically.

Recommended references for further reading appear at the end of each chapter. These are generally well selected, and tend to be relatively recent; references critical of viewpoints discussed in the text tend to be omitted. It is also rather disconcerting to find that Spence, N. E. Miller, Postman and Estes are each credited with three references, yet their names fail to appear in the index.

Since a large number of important topics is dealt with in about two pages each, it must be concluded that the book is not really very satisfactory for use as an undergraduate text-book.

Nevertheless, the excitement which the book sometimes creates suggests that it could serve as a useful antidote to the usual, tedious American text-book. If only someone could write an exhaustive account of present-day psychology which stimulated the enthusiasm of the reader!

MICHAEL W. EYSENCK

A Laboratory Manual for the Control and Analysis of Behaviour. H. LANE and D. BEM. Belmont, California: Wadsworth. Pp. xv + 283. 34*s*.

The book is based on an elementary laboratory course given at the University of Michigan. It describes eleven experiments; four rat, seven human. The language and general tone are heavily Skinnerian.

This is a laboratory manual and has a relatively small proportion of text. Over half of the book is taken up by the forms and visual materials for the experiments. While this no doubt lightens the student's load, the provision of 72 pages for a total of 72 words in an experiment on mediated generalization appears excessive.

There are useful features, e.g. twenty sides of programmed instruction on logarithms and decibels prior to auditory experiments, and also evaluation forms for each of the experiments. The book is concluded by a course evaluation form with the student, being asked *inter alia*, to 'rate the lecturer's general teaching effectiveness'. I would rate the book's effectiveness as low unless used as originally intended with the apparatus and equipment as described. Lehigh Valley Electronics supply this as a package (less animals, food and living cages) at about £300 per experimenter. They will also supply the book free, upon request, to 'qualified persons'. Any rating of the value of the book will be largely dependent on the rater's rating of the Skinnerian approach.

C. ROBSON

Στεφανος: *Studies in Psychology Presented to Cyril Burt.* Edited by CHARLOTTE BANKS and P. L. BROADHURST. University of London Press, 1965. Pp. 284. 42*s*.

A *Festschrift* dedicated to an eminent psychologist with interests and achievements as wide as Sir Cyril Burt's is inevitably something of a mixed grill. Future psychologists will first and foremost turn to *Stephanos* because they are interested in Burt, and they will find a charming and informative biographical sketch by C. W. Valentine, who as Burt's oldest psychological friend knew him for longer and perhaps better than anyone else. (This incidentally must have been among the last things Valentine wrote. It contains one error. R. S. Woodworth did not hold the same post as Burt at Liverpool. He was not a lecturer under Sherrington, but a research fellow.) They will also find a comprehensive bibliography of Burt's writings brought down to 1965—a useful supplement to Dr Steinberg's earlier annotated bibliography published in the *British Journal of Educational Psychology* in 1951.

The thirteen other contributions fall into two main groups: philosophical and theoretical; empirical and experimental. It is not at all clear how and why the various contributions have been selected by the editors. Several distinguished pupils and colleagues of Burt have been omitted, and not all the fields of enquiry in which Burt himself worked have been represented. There are some strange bedfellows among the chosen. Let's hope they never had to meet in person!

The theoretical papers by Mace, Wilkie, Koestler and Gardner Murphy are marked by a strongly anti-behaviourist tone. The tides are certainly running more and more strongly against behaviourism, and Burt no doubt is rejoicing as he enters his ninth decade. Guildford contributes a useful overview of research on creativity but one can't help wondering whether he is on the right track. The outburst of creative achievement at certain periods of human history (Athens, Florence, Elizabethan England) suggest that incentives and the removal of blocks count much more than the discovery of talent. I am not aware that the Athenians or the Florentines or Elizabethans had any method of talent spotting.

Among the empirical articles of particular interest are Gertrude Keir's report on the Tristan da Cunha children, Dr Charlotte Banks's report on boys in detention centres, and a combined report by Broadhurst and Eysenck on emotionality in the rat which lends support to Burt's theory of 'general emotionality'. The contributions are concluded by a paper by R. B. Cattell on higher-order factor structures.

L. S. HEARNSHAW

The Psychology of Egon Brunswik. Edited by KENNETH R. HAMMOND. New York: Holt, Rinehart and Winston. 1966. Pp. 549. $12.00.

Offered as a memorial volume to Egon Brunswik, this collection of articles opens with a generous encomium by Tolman, and contains some fourteen contributions by psychologists who have been Brunswik's students or colleagues or who acknowledge special indebtedness to his views. There are four selected papers by Brunswik himself, including his joint article with Tolman and some sample passages, translated for this volume, from Brunswik's *Wahrnehmung und Gegenstandswelt*, published in Vienna in 1934. Apart from its status as a tribute, the volume will be found valuable as an introduction to Brunswik which, if not a substitute for his original writings, is in many ways easier to read and more relaxed in its manner of exposition. This introduction is provided especially by the editor's own article (which he claims as 'the only paper ever written in America that claims an open alliance with Brunswik's probabilistic functionalism'), and by the contributions of Hochberg and Leeper, both of which are in a section assigned to 'criticisms'. The editor offers some apology for the inclusion of these articles which 'in most memorial volumes would be in poor taste'. But the apology (as perhaps one might expect from its inclusion) is unnecessary, since the special value of both articles is the extent to which they reassess Brunswik's practice by standards that are largely in accord with his own theoretical approach.

The first fifty pages or so of Hammond's own introductory article can be accepted as, in his own words ,'frankly partisan', and emphasize, with particular reference to Harlow's work on error factors in monkey learning, Brunswik's criticism of 'nomothetic behaviourism' as insufficiently concerned with generalization across situations as well as individuals. The argument here, as in Brunswik, seems well designed to stimulate serious reconsideration of many aspects of psychological experimental method. But finally Hammond admits that 'although Brunswik was able to carry out certain studies involving situational sampling, these experiments evidently failed to convince psychologists that true ecological or situation sampling is feasible', and he concludes that 'at this time the question of feasibility remains open'.

Hammond continues with discussion of forms of multi-correlational analysis in which he has himself taken part, and which can perhaps justly be regarded as more rigorous developments in the direct line of Brunswik's work. Other parts of the volume suggest, however, that the very definiteness of Brunswik's methodological suggestions have, by their difficulties in application, tended rather to withdraw attention from his fundamental insights. While both Hochberg and Leeper show serious concern with the problems of representative design, neither displays any particular committal to Brunswik's correlational analysis.

Hochberg, for example, comments on the inadequacy of 'picture sampling' methods to represent the 'normal' locomotive aspects of size and depth perception which have been emphasized by Gibson. His general point is that 'representative sampling' is no substitute for a theory of what to look for, and that Brunswik's sampling procedure often limited the variables his measures could discover just as surely as the limitation of a univariate design. But Hochberg's conclusion that one of the practicable approaches is by behaviour-defined ecological samples does not seem incompatible with Brunswik's adaptational view of an ecology, as being aim- or achievement-defined.

A similar approach to a workable form of ecological sampling is shown by Leeper, who questions the summary of Brunswik's views in Koch, where Postman and Tolman assert that 'when variables are sampled there are as many universes as variables. But there is only one universe of environmental situations, that is, there is only one ecology for a given organism'. Leeper argues for his own 'impression' that Brunswik felt that an individual might be considered from the standpoint of various ecologies. But, he maintains, the definition or identification of an ecology cannot be prior to knowledge of cause-and-effect relationships; and must depend on earlier partial knowledge to separate off cases that are fundamentally similar in some important ways. And, as for Hochberg, this seems to lead towards the definition of a basis for representative design for which such a term as 'behaviour setting' is perhaps more appropriate than the biologically loaded reference to ecology or natural habitat. For the fact is that, if only for the lack of a super-human standpoint, there are very great difficulties in transferring the very useful notion of an ecological niche to the range of human behaviour.

Both these contributors suggest that effective sampling can be simpler than Brunswik usually suggested: e.g. (Hochberg) that 'testing theories, as distinguished from sampling organismic achievement, requires merely that the test conditions represent the *range* of stimulus variation, *not* the shape of its distribution'. This is the tendency of many of Brunswik's experiments, even though Brunswik continued to regret the failure to achieve full representation, Nevertheless, so far as any case is granted for probabilistic learning, the argument for approximately representative design in learning, as against final test situations, is a strong one, and these two types of experiment do not seem as clearly differentiated either in Brunswik or in these papers as they might be.

Hochberg mentions driving ability in traffic as a possible field of sampling procedure. Experiments in this field have recently taken to the open road, certainly a more realistic but also a less controlled situation. A theoretical consideration of the criteria by which such experiments are acceptable might well be a starting point for situational definition and support Brunswik's surmise that such applications will first become standard in 'a psychology eager to do work of practical significance without having to wait for sanctioned academic methods to catch up with their problems'.

A. C. STANILAND

Perception: The Basic Process in Cognitive Development. By RONALD H. FORGUS. London: McGraw-Hill. 1966. Pp. x+402. 72*s*.

Despite the recognized importance of all aspects of perception in modern psychology, there are still relatively few books devoted exclusively to this area. Almost all of them are valuable contributions to teaching or research, which is something that cannot be said for some of the books in areas that are more thoroughly represented. This book is primarily concerned with visual perception and is chiefly intended for undergraduates. It differs from other text-books on visual perception in two ways, the developmental presentation of the chapters and the incorporation of material concerned with various aspects of thinking.

In the first two chapters, Forgus attempts to show visual perception to be a process of information extraction from the environment. Perception is represented as a hierarchy of 'subtasks' beginning with the simplest detection of stimuli, passing through the stages of figural unity, identification, etc., and ending in social perception and the higher thought processes. The succeeding chapters follow this pattern. After presenting some basic methodology and elementary psychophysics in chapters 3 and 4, the next two chapters deal with brightness constancy and with size and shape constancy. Chapters 7 and 8 are concerned with figure and pattern perception, incidentally illuminated by an elementary and useful description of information theory. The following two chapters deal with perceptual learning and space perception respectively. The section on perception ends with two chapters (11 and 12) on motion perception, time perception and social perception in general.

The remaining part of the book is unusual for a text on visual perception in that it consists of four chapters dealing progressively with psychological sets in thinking, concept formation and language, problem solving and finally productive and creative thinking. Although Forgus states clearly that his book is a teaching instrument and does not attempt to be argumentative, his initial thesis that '... perception is the superset of information extraction, with learning and thinking as subsets', is not at all adequately demonstrated in any systematic or thorough way. The last four chapters on thinking seem to be dealing with complex cognitive thought processes and not with a continuation or subset of perception.

Another difficulty that one anticipates will be experienced by the student is the author's failure to distinguish clearly between the development of perception during the lifetime of the organism on the one hand and the presumed stages during any particular act of perception on the other. Again, his presentation at the beginning of the book is confusing at times when he refers to the concept of 'set'. Sometimes he uses this term in the psychological sense and sometimes in its logical sense without distinguishing clearly between them.

Apart from these objections, this is a useful book. In general, it is clearly presented and the author does not hesitate to use material both from early sources and the most recent literature, ranging from the Würzburg school to information theory and ethology. B. E. SINGER

Human Spatial Orientation. By I. P. HOWARD and W. B. TEMPLETON. London: Wiley. 1966. Pp. 533. 84*s*.

This is a timely, stimulating and enjoyable book. Having recognized the need for a volume assembling all the evidence relevant to their central topic, the authors have fulfilled this need admirably.

'This book is about those aspects of human behaviour which are determined by the angular position of the body (or head) in relation to any stable external reference system.' The allied topic of judging the inclination of lines to gravity is specifically included. In the first part of the book the authors review the present state of knowledge in the four most relevant modalities: vision, especially visual direction; kinaesthesis; the vestibular apparatus (labyrinths); and auditory localization. The second part of the book is devoted to a comprehensive coverage of the central topic. It surveys work on orientation to gravity with reference to visual cues, postural cues, adaptation effects, and interactions among these. It reviews current knowledge on geographical orientation and on egocentric orientation. The effects of shape on perceived orientation and the effects of orientation on perceived shape are discussed, and developmental aspects of the relationship between orientation and shape are described. Current knowledge on sensorimotor and intersensory localization is assembled; and a further chapter details the behavioural consequences of rotations and displacements of the optical array. A final chapter reviews the generally accessible knowledge on orientation in the weightless state.

The authors have assembled and linked together psychological, physical and physiological data, while maintaining a psychological viewpoint. They make the bold claim that all the chapters in the second part of the book except the last one are comprehensive; on the whole their claim is justified. As a result, one of the most valuable features of this book is its list of references—80 pages of them. Their value is enhanced by thorough cross-referencing to the text. The authors have organized their material well, and each chapter is prefaced by a list of headings which clearly define its range and contents. Good organization of material has helped the authors to specify exactly those issues which have yet to be tackled or which remain unresolved, and helps the reader to see these in their correct context. The authors write clearly, without imposing a literary style which is obtrusive in any way. On occasion they have been tempted to reformulate in their own terms the basic tenets of others whose clarity of expression leaves something to be desired; not surprisingly such a hazardous enterprise is not always vindicated by the results.

This book is wide-ranging and thorough but it does more than present a summary of facts. The authors persistently and engagingly scrutinize experimental methods, question interpretations of results, point out unrecognized implications of findings, note incompatibilities and inconsistencies and advance alternative explanations. Some of their comments, for example in regard to the role of centrifuges, may be contentious but most of their criticisms are apt and welcome. They are aimed at inadequacies of experimental design, at bias in interpreting results, and, perhaps most often, at the kind of blinkered approach which narrows the area of search for alternative explanations. Although most criticisms are directed against slovenliness in the conception, conduct, or interpretation of experiments, this does not imply that the authors are indifferent to theoretical views. Their willingness to criticize unproductive theoretical models is exemplified by their condemnation of sensoritonic theory at every opportunity.

As the first comprehensive survey of human spatial orientation, this book seems an essential library purchase. Its qualities as a source book, its thoroughness, and the critical approach adopted by the authors earn the recommendation that all whose work or interests touch on human spatial orientation would do well to buy it. V. DAVID HOPKIN

Psychology, 5th edition. By NORMAN L. MUNN. Boston: Houghton Mifflin Company. 1966. Pp. xv + 716. $8.50.

Psychology, 2nd edition. By HENRY CLAY LINDGREN, DONN BYRNE, LEWIS PETRINOVICH. New York: Wiley, 1966. Pp. xv + 560. 60*s*.

The main difficulty in reviewing an introductory text-book is to arrive at the criteria by which it is to be judged. This difficulty is brought even more sharply into focus when the reviewer is simultaneously faced with two such books. Here, for instance, the two books are both in a sense introductory, both in a sense cover much the same ground and yet seem to this reviewer to operate at so different a level that it is scarcely meaningful to compare them.

To begin with the similarities, the scope of the two books is superficially much the same. Both cover basic processes such as learning and perception, memory, emotion and motivation, individual differences, developmental and social psychology. Both at some stage provide an outline of biological and physiological information relevant to an understanding of behaviour. From the standpoint of basic content alone perhaps the main difference between them is that Munn pays more attention to biological factors and to thinking and language, while Lindgren, Byrne and Petrinovich seem more concerned with social issues and broad problems of applied psychology.

Within this, however, there is a much more fundamental difference, a difference in terms of levels of analysis, exposition and explanation. It is perhaps symptomatic that in the case of Munn the reader is constantly aware of the essentially biological orientation of the approach, and the willingness to pursue complex issues—while in the case of Lindgren *et al.* the orientation is much more blatantly towards showing and even justifying the relevance of psychology to everyday life.

The effect of this difference is profound. Munn, from the very beginning, makes an attempt to set out and explain what might be called 'the psychological approach' to the study of behaviour. He examines not only what psychologists do and how they do it, but *why* certain problems are important, why certain ways of tackling them were considered, used, modified or discarded. In other words Munn presents his reader not simply with the subject-matter of present-day psychology, but also with an exposition of the logic of the methods and approaches which have been adopted. It is thus not only concerned with the subject matter, but with the kind of thinking which underlies the intellectual attack on it.

By contrast Lindgren, Byrne and Petrinovich's introduction is essentially a survey of the kind of work carried out under the general heading of 'Psychology'. Though specific methodological problems are considered, the reader is largely left to himself to work out the reasoning behind the approaches and theories adopted. On the other hand this book consistently attempts to relate psychological findings to the everyday problems to which they might possibly be relevant. Simultaneously it makes considerable use of essentially anecdotal evidence, at least in the form of invoking the reader's own experience to emphasize a point or clarify an argument.

From an academic standpoint this reviewer is in no doubt that he would prefer his students to read Munn rather than Lindgren *et al.* But this is a function of his students, all of whom are potential specialists in psychology. Yet there are other audiences: students and people generally who may want to know more about psychology—not to pursue it in depth but rather in terms of everyday problems which they encounter and feel worth thinking about. At this much more general level Lindgren, Byrne and Petrinovich have something genuine to offer: for precisely because their *Psychology* lacks depth compared with Munn's it also demands much less commitment on the part of the less ambitious reader. But there is room both for the more academically oriented approach of Munn and the competent but everyday orientation of Lindgren, Byrne and Petrinovich.

R. P. KELVIN

Personality: An Objective Approach. By IRWIN G. SARASON. London: Wiley, 1966. xvi + 670. 62*s*.

Irwin Sarason's book has five parts: views of personality; personality assessment; the experimental study of personality; personality development and deviant behaviour.

'Views of personality' are mainly those of Freud, Jung, Adler, Fromm, Horney, Sullivan, Hartmann, Dollard and Miller, Rogers, Gordon Allport and Lewin. These stalwarts have, of course, been expounded with excruciating fairness many times in the past. Sarason certainly does it well without adding much that is different from, say, Hall and Lindzey, except perhaps for more on Hartmann and the ego psychologists. This reiteration, though somewhat tedious for reviewers, can, however, be justified in the total context.

'Personality assessment' is in much the same case, being very similar to parts of Anastasi, Cronbach or Vernon. At the later date of this publication we might have been given more about G. A. Kelly and Holtzman's ink-blot technique. The fact that Cattell is not included in the first part and appears very briefly elsewhere reflects the rather scant attention given to the statistical treatment of personality measures, despite the author's recognition that the experimental method cannot be reckoned superior to other methods in all circumstances. A brief description of a study using multivariate analysis could have been included with advantage; on the other hand, even very brief sections on genetic, neurological and physiological influences on behaviour are very welcome in an American book of this kind.

With 'The experimental study of personality' we come to what is perhaps the best part of the book. The material selected is probably less widely known and is sufficiently small to allow the author room to describe projects interestingly and to illustrate effectively points he wishes to make. An exception, which supports the desirability of the rule, is the brief account of a study by Scodel and Mussen in which pairs of high and low authoritarians had to fill out an attitude questionnaire in the way each thought his partner would do so. We are told that highs rated their partners high and were, therefore, 'wrong'; lows rated their partners high and were, therefore, 'right'. This is only true in a misleading sense. Both groups, in fact, rated their partners high. The design is inadequate for demonstrating accuracy of interpersonal perception and required high–high and low–low pairings.

'Personality development' appears to be sound and interesting. Here, as elsewhere, the author pulls out the questions, as well as the answers, which need pursuing. He is fully alive to the paramount importance and the great difficulty of longitudinal as opposed to cross-sectional studies. An even more detailed spelling out of the methodological differences and the danger of inferring longitudinal conclusions from cross-sectional results would be worth while.

In 'Deviant behaviour' diagnostic issues (such as clinical *v.* statistical prediction) treatment issues (such as psychotherapy *v.* behaviour therapy) and aetiological issues (such as environmental *v.* innate influences) have the steam taken out of them. Studies of psychotherapy now tend to include the therapist as well as the patient and there are other welcome signs of methodological sophistication with which behaviour therapists will need to catch up.

Sometimes, as with so many similar American texts, the therapeutic handling of some of our professional conflicts leaves one hankering for the occasional more rumbustuous criticism. The much publicized work of Kallmann, for example, should no longer escape unscathed.

In the preface Sarason states that the book contains the material he covers with undergraduates at Washington University. Students from elsewhere can now share their good fortune.

G. A. FOULDS

Brit. J. Psychol. (1966), 57, 3 *and* 4, *pp.* 503–512
Printed in Great Britain

OTHER PUBLICATIONS RECEIVED

Abraham, H. C. & Freud, E. L. (eds.). *A Psycho-analytic Dialogue. The Letters of Sigmund Freud and Karl Abraham, 1907–1926.* The International Psycho-analytical Library, no. 68. London: Pergamon. 1965. Pp. xvii + 406. 63*s*.

Abse, D. W. *Hysteria and Related Mental Disorders.* Bristol: Wright and Sons. 1966. Pp. xi + 292. 42*s*.

Akos, K. & Akos, M. *The Critical Flicker Frequency Series Effect.* Budapest: Akademiai Kiado. 1966. Pp. 245. $7.50.

Altman, J. *Organic Foundations of Animal Behavior.* New York: Holt, Rinehart and Winston. 1966. Pp. xiv + 530. $12.95.

Anastasi, A. *Individual Differences.* London: Wiley, 1965. Pp. xiii + 301. Cloth 40*s*. Paper 25*s*.

Andor, L. E. *Aptitudes and Abilities of the Black Man in Sub-Saharan Africa 1784–1963.* Johannesburg: National Institute for Personnel Research. 1966. Pp. vii + 174. No Price given.

Andrews, G. & Harris, M. *The Syndrome of Stuttering.* London: Developmental Medicine and Child Neurology. 1964. Pp. 191. 40*s*.

Abell, W. *The Collective Dream in Art.* New York: Schocken. 1966. xv + 378. 26*s*.

Aspects of Anxiety. Philadelphia: Lippincott. 1966. Pp. 80. 30*s*.

Assogioli, R. *Psychosynthesis: A Manual of Principles and Techniques.* New York: Hobbs Dorman. 1965. Pp. 323. $7.50.

Atkinson, J. W. & Feather, N. T. *A Theory of Achievement Motivation.* London: Wiley. 1966. Pp. x + 392. 87*s*.

Atkinson, R. C. (ed.). *Studies in Mathematical Psychology.* London: Oxford University Press. 1964. Pp. vii + 414. 92*s*.

Balint, M. *et al. A Study of Doctors.* London: Tavistock. 1966. Pp. xii + 146. 28*s*.

Ban, T. A. *Conditioning and Psychiatry.* London: George Allen and Unwin. 1966. Pp. ix + 244. 40*s*.

Banerjee, J. C. *Fundamentals of Modern Psychology. Part I.* London: Asia Publishing House. 1966. Pp. xxx + 317. 35*s*.

Barker, R. G. & Wright, H. F. *One Boy's Day: A Specimen Record of Behaviour.* London: Archon. 1966. Pp. x + 435. 87*s*. 6*d*.

Barrett, C. (ed.). *Wittgenstein.* Oxford: Blackwell. 1966. Pp. 72. 12*s*. 6*d*.

Beach, F. A. *Sex and Behaviour.* London: Wiley. 1965. Pp. xvi + 592. 75*s*.

Bennet, E. A. *What Jung Really Said.* London: Macdonald. 1966. Pp. 185. 15*s*.

Bergmann, T. *Children in the Hospital.* New York: International Universities Press. 1966. Pp. 162. 27*s*.

Biggar, J. *Psychotherapy and Child Dvelopment.* London: Tavistock 1966. Pp. xi + 142. 25*s*.

Bilodeau, E. A. *Acquisition of Skill.* New York and London: Academic Press. 1966. Pp. xii + 538. 100*s*.

Bindra, D. & Stewart, J. *Motivation.* Harmondsworth: Penguin. 1966. Pp. 352. 8*s*. 6*d*.

Biran, S. *Die auberpsychologischen Voraussetzungen der tiefenpsychologie.* Munchen, Basle: Verlag. 1966. Pp. 180. DM 11.

Black, S. *Man and Motor Cars. An Ergonomic Study.* London: Secker and Warburg. 1966. Pp. 373. 63*s*.

Blank, L. *Psychological Evaluation in Psychotherapy.* London: Allen and Unwin. 1966. Pp. xii + 364. 70*s*.

Blommers, P. & Lindquist, E. F. *Elementary Statistical Methods in Psychology and Education.* London: University of London Press. 1966. Pp. xv + 528. 50*s*.

Blommers, P. & Lindquist, E. F. *Manual for Elementary Statistical Methods in Psychology and Education*. London: University of London Press. 1966. Pp. 248. 17*s*. 6*d*.

Blum, G. A. *Psychodynamics: The Science of Unconscious Mental Forces*. California: Wadsworth. 1966. Pp. x + 22. 12*s*.

Bolland, J. & Sandler, J. *The Hampstead Psychoanalytic Index*. Monograph Series for the Psychoanalytic Study of the Child. London: Hogarth. 1967. Pp. xiii + 205. 35*s*.

Borger, R. & Seaborne, A. E. M. *The Psychology of Learning*. Harmondsworth: Penguin. 1966. Pp. 5*s*.

Boorer, D. & Boorer, H. *An Introduction to Psychiatric Nursing*. London: Pergamon. 1966. Pp. ix + 175. 12*s*. 6*d*.

Bottomore, T. B. *Classes in Modern Society*. London: Allen and Unwin. 1966. Pp. 82. Cloth 16*s*. Paper 9*s*. 6*d*.

Bowlby, J. *Maternal Care and Mental Health* and *Deprivation of Maternal Care*. New York: Schocken. 1966. Pp. 357. 22*s*.

Brim, O. C. & Wheeler, S. *Socialization after Childhood. Two Essays*. London Wiley. 1966. Pp. vi + 116. Cloth 38*s*. Paper 17*s*.

Britt, S. H. *Consumer Behaviour and the Behavioural Sciences*. London: Wiley. 1966. Pp. xxxii + 592. 87*s*.

Brown, J. M., Berrien, F. K. & Russell, D. L. *Applied Psychology*. London: Collier Macmillan. 1966. Pp. xiii + 639. 70*s*.

Bruner, J. S. *Toward a Theory of Instruction*. London: Oxford University Press. 1966. Pp. x + 176. 32*s*.

Bruner, J. S., Olver, R. R. & Greenfield, P. M. *Studies in Cognitive Growth*. London: Wiley. 1966. Pp. xviii + 343. 60*s*.

Burns, T. & Stalker, G. M. *The Management of Innovation*. London: Tavistock. 1966. Pp. xii + 269. 35*s*.

Brown-Grant, K. & Gross, B. A. (eds.). Recent Studies on the Hypothalamus. *Brit. Med. Bull.*, Vol. **22**. No. 3. September 1966. London: British Council. 1966. Pp. 195–284. 30*s*.

Buros, O. K. (ed.). *The Sixth Mental Measurements Yearbook*. New Jersey: Gryphon Press. 1965. Pp. xxxvi + 1714. $32.50.

Buss, A. H. *Psychopathology*. London: Wiley. 1966. Pp. xi + 483. 60*s*.

Carter, C. H. *Handbook of Mental Retardation Syndromes*. Springfield, Ill.: Thomas. 1966. Pp. xii + 168. $8.00.

Chapanis, A. *Man-Machine Engineering*. London: Tavistock. 1966. Pp. vii + 134. 21*s*.

Chapanis, A. *Research Techniques in Human Engineering*. Baltimore: Johns Hopkins. London: Oxford University Press. 1966. Pp. xii + 316. 20*s*.

Chapman, A. J. *The Management of Emotional Problems of Children and Adolescents*. Philadelphia: Lippincott. 1965. Pp. xi + 316. 80*s*.

Chertok, L. *Hypnosis*. Oxford: Pergamon. 1966. Pp. xv + 176. 45*s*.

Chesser, E. *Living with Suicide*. London: Hutchinson. 1967. Pp. 176. 30*s*.

Chiselli, E. E. *The Validity of Occupational Aptitude Tests*. London: Wiley. 1966. Pp. vii + 155. 60*s*.

Church, J. *Three Babies: Biographies of Cognitive Development*. New York: Random House. 1966. Pp. ix + 323. $2.95.

Cicala, G. A. *Animal Drives*. Princeton: Van Nostrand. 1965. Pp. vi + 266. 15*s*. 6*d*.

Clarke, A. M. & Clarke, A. D. B. *Mental Deficiency: The Changing Outlook*. London: Methuen. 1966. Pp. xxi + 596. 63*s*.

Clausen, J. *Ability Structures and Subgroups in Mental Retardation*. London: Macmillan. 1966. Pp. vii + 208. 80*s*.

Clyne, M. B. *Absent. School Refusal as an Expression of Disturbed Family Relationships*. London: Tavistock. 1966. Pp. xi + 269. 38*s*.

Cohen, J. *A New Introduction to Psychology.* London: Allen and Unwin. 1966. Pp. 220. Cloth 25*s.* Paper 12*s.* 6*d.*

Cole, J. D. & Wittenborn, J. R. *Pharmacotherapy of Depression.* Springfield, Ill.: Thomas. 1966. Pp. xi + 189. $8.50.

Cole, W. G. *Sex in Christianity and Psychoanalysis.* London: Oxford University Press. 1967. Pp. xvi + 329. 12*s.* 6*d.*

Conrad, H. S. *Studies in Human Development.* New York: Appleton-Century-Crofts. 1966. Pp. xiv + 392. $6.00.

Cooper, D. *Psychiatry and Anti-Psychiatry.* London: Tavistock. 1967. Pp. xiii + 128.

Coopersmith, S. (ed.). *Frontiers of Psychological Research. Readings from Scientific American.* London: Freeman. 1966. Pp. xiii + 322. Paper 36*s.*

Costello, C. G. *Psychology for Psychiatrists.* Oxford: Pergamon. 1966. Pp. x + 328. 25*s.*

Courts, F. A. *Psychological Statistics.* Homewood, Ill.: Dorsey. 1966. Pp. x + 349. $7.50.

Craft, M. (ed.). *Psychopathic Disorders and their Assessment.* Oxford: Pergamon. 1966. Pp. x + 320. 50*s.*

Crane, E. *Marketing Communications.* London: Wiley. 1965. Pp. xi + 569. 68*s.*

Creelman, M. B. *The Experimental Investigation of Meaning. A Review of the Literature.* New York: Springer. 1966. Pp. xi + 228. $6.75.

Cunnison, S. *Wages and Work Allocation.* London: Tavistock. 1966. Pp. xxxvi + 291. 42*s.*

Decarie, T. G. *Intelligence and Affectivity in Early Childhood.* New York: International Universities Press. 1966 Pp. xvi + 230. 45*s.*

Dana, R. H. *Foundations of Clinical Psychology.* London: Van Nostrand. 1966. Pp. viii + 322. 54*s.*

Deese, J. *The Structure of Language and Thought.* Baltimore: Johns Hopkins. London: Oxford University Press. 1966. Pp. xiii + 216. 52*s.*

Dennis, W. *Group Values Through Children's Drawings.* London: Wiley. 1966. Pp. xiii + 211. 53*s.*

Deutsch, J. *Neuroses and Character Types.* London: Hogarth. 1966. Pp. xii + 388. 42*s.*

Diamond, S. *Information and Error.* London: Basic Books. 1966. Pp. xii + 307. 27*s.* 6*d.*

Douvan, E. & Adelson, J. *The Adolescent Experience.* London: Wiley. 1966. Pp. xii + 472. 60*s.*

Downes, D. M. *The Delinquent Solution. A Study in Subcultural Theory.* London: Routledge and Kegan Paul. 1966. Pp. xi + 284. 42*s.*

Duker, S. *Listening: Readings.* New York and London: Scarecrow Press. 1966. Pp. 474. 110*s.*

Eccles, J. *The Brain and the Unity of Conscious Experience.* London: Cambridge University Press. 1965. Pp. 45. 4*s.* 6*d.*

Eckley, P. A. & Janecek, C. B. *Against the Tide.* New York: Exposition Press. 1966. Pp. 166. $5.00.

Erikson, E. H. *Insight and Responsibility.* London: Faber and Faber. 1966. Pp. 256. 30*s.*

Ehrenstein, W. *Probleme des hoheren Seelenlebens.* Munchen, Basle: Verlag. 1966. Pp. 501. No price given.

Evans, C. R. & Robertson, A. D. J. *Brain Physiology and Psychology. Key Papers.* London: Butterworths. 1966. Pp. viii + 280. 32*s.* 6*d.*

Fairbanks, G. *Experimental Phonetics: Selected Articles.* Urbana, Ill.: University of Illinois Press. 1966. Pp. 274. 49*s.*

Farber, B. (ed.). *Kinship and Family Organisation.* London: Wiley. 1966. Pp. ix + 460. Cloth 64*s.* Paper 34*s.*

Farson, R. *et al. Science and Human Affairs.* Palo Alto: California. 1966. Pp. viii + 214. $4.95.

Fechner, G. *Elements of Psychophysics.* New York: Holt, Rinehart and Winston. Pp. xxxi + 286. $6.00.

Fensham, P. J. & Hooper, D. *The Dynamics of a Changing Technology: A Case Study in Textile Manufacturing.* London: Tavistock. 1965. Pp. xvi + 248. 35*s.*

Ferdinand, T. N. *Typologies of Delinquency.* New York: Random House. 1966. Pp. x + 246. $2.25.

Ferguson, G. A. *Non-parametric Trend Analysis.* Leicester: University Press. 1965. Pp. v + 62. 15s.

Fernald, L. D. *Experiments and Studies in General Psychology.* London: Harrap. 1965. Pp. xii + 200. 15s.

Fernald, L. D. & Fernald, P. S. *Overview of General Psychology: A basic Program.* London: Harrap. 1966. Pp. ix + 354. 21s.

Fitts, W. H. *The Experience of Psychotherapy.* Princeton: Van Nostrand. 1965. Pp. 188. 15s. 6d.

Foss, B. M. (ed.). *New Horizons in Psychology.* Harmondsworth: Penguin. 1966. Pp. 448. 7s. 6d.

Frankenstein, C. *The Roots of the Ego.* London: Livingstone. 1966. Pp. xi + 286. 62s.

Freud, A. *Normality and Pathology in Childhood.* London: Hogarth. 1966. Pp. xii + 274. 35s.

Friedman, A. A. *et al. Psychotherapy for the Whole Family.* New York: Springer. 1966. Pp. vii + 354. $7.50.

Frosche, J. & Ross, N. *The Annual Survey of Psychoanalysis.* Vol. 8. London: Hogarth. 1965. Pp. xii + 371. 100s.

Gabe, M. *Neurosecretions.* Oxford: Pergamon, 1966. Pp. xiv + 872. 200s.

Gaito, J. *Macromolecules and Behaviour.* Amsterdam: North Holland. 1966. Pp. ix + 197. 58s.

Galanter, E. *Textbook of Elementary Psychology.* San Francisco: Holden-Day. 1966. Pp. xiii + 419. $8.75.

Gavron, H. *The Captive Wife.* London: Routledge and Kegan Paul. 1966. Pp. xiv + 176. 25s.

Giallombardo, R. *Society of Women. A Study of a Women's Prison.* London: Wiley. 1966. Pp. vii + 244. 49s.

Gilberstadt, H. & Duker, J. *A Handbook for Clinical and Actuarial MMPI Interpretation.* Philadelphia and London: Saunders. 1965. Pp. x + 134. 38s.

Gill, M. M. & Brenman, M. *Hypnosis and Related States.* London: Wiley. 1966. Pp. xxiv + 404. 17s.

Goldstein, A. P., Heller, K. & Sechrest, L. B. *Psychotherapy and the Psychology of Behaviour Change.* New York and London: Wiley. 1966. Pp. ix + 472. 68s.

Gordon, I. J. *Studying the Child in School.* London: Wiley. 1966. Pp. ix + 145. Cloth 38s. Paper: 23s.

Goss, A. E. & Nodine, C. F. *Paired-Associates Learning: The Role of Meaningfulness, Similarity and Familiarization.* New York: Academic Press. 1966. Pp. ix + 358. $13.50.

Granit, R. (ed.). *Nobel Symposium: I. Muscular Afferents and Motor Control.* London: Wiley. 1966. Pp. 466. 150s.

Greco, R. S. & Pittenger, R. A. *One Man's Practice.* London: Tavistock. 1966. Pp. xviii + 124. 25s.

Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics.* London: Wiley. 1966. Pp. xi + 455. 104s.

Guilford, J. P. *Fields of Psychology.* London: Van Nostrand. 1966. Pp. x + 350. 72s.

Hall, C. S. & Van de Castle, R. L. *The Content of Dream Analysis.* New York: Appleton-Century-Crofts. 1966. Pp. xiv + 320. $7.00.

Halloran, J. D. *Effects of Mass Communication.* Leicester: Leicester University Press. 1964. Pp. 83. 7s. 6d.

Harary, F. *et al. Structural Models.* London: Wiley. 1965. Pp. viii + 416. 75s.

Hatwell, Y. *Privation Sensorielle et Intelligence.* Paris: Presses Universitaires de France. 1966. Pp. 232. 16F.

Haward, L. R. C. *Reading Comprehension Test for Personnel Selection.* London: University of London Press. 1966. Manual 4s. Answer Paper 6d.

Hebb, D. O. *A Textbook of Psychology.* 2nd Edition. London: Saunders. 1966. Pp. xi + 353. 40s.

Heckel, R. V. & Peacock, L. J. *Textbook of General Psychology*. London: Kimpton. Saint Louis: Mosby. 1966. Pp. xii + 480. 63*s*.

Helmstadter, G. C. *Principles of Psychological Measurement*. London: Methuen. 1966. Pp. xx + 248. 36*s*.

Hilgard, E. R. *Theories of Learning and Instruction*. Chicago: University of Chicago Press. 1965. Pp. xi + 430. 41*s*.

Hiltmann, H. *Kompendium der psychodiagnostischen Tests*. Bern: Huber. 1966. Pp. 304. Fr./DM 22.

Hobhouse, L. T. *Sociology and Philosophy. A Centenary Collection of Essays and Articles*. London: Bell. 1966. Pp. xxvii + 340. 45*s*.

Hoffer, A. & Osmond, H. *How to Live with Schizophrenia*. London: Johnson. 1966. Pp. ix + 172. 30*s*.

Holland, G. *Fundamentals of Psychotherapy*. New York: Holt, Rinehart and Winston. 1965. Pp. xi + 308. $6.95.

Honig, W. (ed.) *Operant Behavior: Areas of Research and Behavior*. New York: Appleton-Century-Crofts. 1966. Pp. xi + 865. $12.00.

Horrocks, J. E. *Assessment of Behaviour*. Columbia, Merrill. 1966. Pp. xv + 736. No price given.

Hutt, M. L. & Gibby, R. G. *The Mentally Retarded Child*. Boston: Allyn & Bacon. 1965. Pp. xiv + 441. $7.95.

Hutt, M. L., Isaacson, R. L. & Blum, M. L. *Psychology: The Science of Interpersonal Behaviour*. New York: Harper and Row. 1966. Pp. xii + 430. 30*s*.

Iverson, L. L. *The Uptake and Storage of Noradrenaline in Sympathetic Nerves*. Cambridge: University Press. 1967. Pp. xiii + 232. 57*s*. 6*d*.

Jahoda, M. & Warren, N. *Attitudes*. Harmondsworth: Penguin. 1966. Pp. 376. 8*s*. 6*d*.

Jordan, T. E. *The Mentally Retarded*. London: Prentice Hall. 1966. Pp. xl + 451. 60*s*.

Jung, C. G. *Psychiatrische Studien*. Zurich and Stuttgart: Verlag. 1966. Pp. xii + 281. S.Fr. 32.

Jung, J. & Bailey, J. *Contemporary Psychology Experiments*. London: Wiley. 1966. Pp. vii + 158. 23*s*.

Junge, K. *Some Problems of Measurements in Psychophysics*. 1966. Pp. 81. N.Kr. 18.

Kalish, R. A. *The Psychology of Human Behaviour*. Belmont, California: Wadsworth. London: Prentice Hall. 1967. Pp. viii + 529. 56*s*.

Kaplan, M. (ed.). *Essential Works of Pavlov*. New York: Bantam. 1966. Pp. ix + 392. $1.25.

Katz, D. & Kahn, R. L. *The Social Psychology of Organisations*. London: Wiley. 1966. Pp. viii + 498. 64*s*.

Kausler, D. H. *Readings in Verbal Learning*. London: Wiley. 1966. Pp. xii + 578. 60*s*.

Kelman, H. (ed.). *International Behaviour: A Social Psychological Analysis*. New York: Holt. 1965. Pp. xiv + 626. $12.95.

Ketchum, J. D. *Ruhleben. A Prison Camp Society*. Toronto: Toronto University Press. London: Oxford University Press. 1966. Pp. xiii + 397. 48*s*.

Kinsey, B. A. *The Female Alcoholic*. Springfield, Ill.: Thomas. 1966. Pp. viii + 233. $8.75.

Klare, H. J. *Changing Concepts of Crime and its Treatment*. Oxford: Pergamon. 1966. Pp. ix + 186. 50*s*.

Klein, M. *et al*. *Developments de la psychoanalyse*. Paris: Presses Universitaires de France. 1966. Pp. 343. F 24.

Klemer, R. H. *Counselling in Marital and Sexual Problems*. Baltimore: Williams and Wilkins. 1965. Pp. x + 310. 78*s*.

Kneller, G. F. *Logic and Language of Education*. London: Wiley. 1966. Pp. viii + 242. Cloth 45*s*. Paper 27*s*.

Kolson, C. J. & Kaluger, G. *Clinical Aspects of Remedial Reading*. Springfield, Ill.: Thomas. 1963. Pp. xii + 146. $5.75.

Krafft-Ebing, R. von. *Psychopathia Sexualis*. London: Staples. 1965. Pp. xvii + 434. 50*s*.

Kreitler, S. *Symbolschopfung und Symbolerfassung*. Munchen, Basle; Verlag. 1966. Pp. 213. DM. 20.

Kuhn, A. *The Study of Society: A Multidisciplinary Approach*. London: Tavistock. 1966. Pp. xviii + 810. 63*s*.

Laing, R. D. *The Politics of Experience* and *The Bird of Paradise*. Harmondsworth: Penguin, 1967. Pp. 156. 3*s*. 6*d*.

Laing, R. D., Phillipson, H. & Lee, A. R. *Interpersonal Perception*. London: Tavistock. 1966. Pp. vii + 179. 35*s*.

Lamb, W. *Posture and Gesture*. London: Duckworth. 1965. Pp. 190. 25*s*.

Lampl de Groot, J. *The Development of the Mind*. London: Hogarth. 1966. Pp. x + 391. 45*s*.

Lidz, T., Fleck, S. & Cornelison, A. R. *Schizophrenia and the Family*. New York: International Universities Press. 1966. Pp. 478. 90*s*.

Lifton, W. M. *Working with Groups*. London: Wiley. 1966. Pp. xii + 288. 53*s*.

Litman, R. E. *Psychoanalysis in the Americas*. New York: International Universities Press. 1966. Pp. xi + 315. 68*s*.

Lorenz, K. *Evolution and Modification of Behavior*. London: Methuen. 1966. Pp. 121. 25*s*.

Loring, J. (ed.). *Teaching the Cerebral Palsied Child*. London: Spastics Society and Heinemann. 1965. Pp. 231. 15*s*.

Lorr, M. (ed.). *Explorations in Typing Psychotics*. Oxford: Pergamon. 1966. Pp. 241. 60*s*.

Loveday, R. *Statistics: A First Course*. London: Cambridge University Press. 1966. Pp. xii + 129. 9*s*. 6*d*.

Lowenstein, O. *The Senses*. Harmondsworth: Penguin. 1966. Pp. 214. 5*s*.

Luria, A. R. *Higher Cortical Functions in Man*. London: Tavistock. 1966. Pp. xvi + 514. 126*s*.

Luria, A. R. *Human Brain and Psychological Processes*. New York and London: Harper and Row. 1966. Pp. xix + 587. 115*s*.

McGrath, J. E. & Altman, I. *Small Group Research: A Synthesis and Critique of the Field*. London: Holt, Rinehart and Winston. 1966. Pp. viii + 601. 100*s*.

Maher, B. A. *Progress in Experimental Personality Research*. New York and London: Academic Press. 1966. Pp. x + 332. $9.50.

McKeachie, W. J. & Doyle, C. L. *Psychology*. Reading, Massachusetts: Addison Wesley. 1966. Pp. xv + 704. 49*s*.

Mackenzie, N. (ed.). *A Guide to the Social Sciences*. London: Weidenfeld and Nicolson. 1966. Pp. 268. Cloth 42*s*. Paper 18*s*.

McLaughlin, B. E. *Long-Term Results of Psychiatric Out-patient Treatment*. Springfield, Ill.: Thomas. 1966. Pp. vii + 80. $4.75.

MacLeod, A. W. *Recidivism: A Deficiency Disease*. London: Oxford University Press. 1966. Pp. 131. 36*s*.

Macnamara, J. *Bilingualism and Primary Education: A Study of Irish Experience*. Edinburgh: University of Edinburgh Press. 1966. Pp. viii + 172. 35*s*.

Maier, N. R. F. *Principles of Human Relations*. London: Wiley. 1966. Pp. ix + 474. 21*s*.

Mannin, E. *Loneliness*. London: Hutchinson. 1966. Pp. 192. 30*s*.

Masland, R. L. *et al*. *Mental Subnormality*. New York: Basic Books. 1966. Pp. 442. 27*s*. 6*d*.

May, R. *Psychology and the Human Dilemma*. Princeton: Van Nostrand. 1967. Pp. 221. $5.95.

Miller, G. A. *Psychology: The Science of Mental Life*. Harmondsworth: Penguin. 1966. Pp. 415. 7*s*. 6*d*.

Moles, A. (Trans. J. E. Cohen). *Information Theory: An Esthetic Perception*. Urbana, Ill.: University of Illinois Press. 1966. Pp. 312. $7.50.

Money, J. *A Standardised Road-Map of Direction Sense*. Baltimore: Johns Hopkins Press. London: Oxford University Press. 1965. Pp. 32. 52*s*.

Moriarty, A. E. *Constancy and I.Q. Change*. Springfield, Ill.: Thomas. 1966. Pp. 222. $8.50.

Morris, D. *The Biology of Art*. London: Methuen. 1966. Pp. 176. 42*s*.

Moustakas, C. *Existential Child Therapy*. London: Basic Books. 1966. Pp. xii + 254. 45*s*.

Murphy, G. *Personality: A Biosocial Approach to Origins and Structure*. London: Basic Books. 1966. Pp. xiv + 999. 27*s*. 6*d*.

Nagera, H. *Early Childhood Disturbances, the Infantile Neurosis, and the Adulthood Disturbances*. London: Hogarth. 1967. Pp. 93. 25*s*.

Nervous and Hormonal Mechanisms of Integration. Symposia of the Society for Experimental Biology. Cambridge: University Press. 1966. Pp. viii + 565. 75*s*.

Neubauer, P. B. *Children in Collectives*. Springfield, Ill.: Thomas. 1965. Pp. xxv + 384. $11.50.

Newman, J. W. (ed.). *On Knowing the Consumer*. New York and London: Wiley. 1966. Pp. viii + 248. 53*s*.

Nisbet, J. D. & Entwistle, N. J. *The Age of Transfer to Secondary Education*. London: University of London Press. 1966. Pp. 112. 12*s*. 6*d*.

Offer, D. & Sabshin, M. *Normality*. London: Basic Books. 1966. Pp. xii + 254. 35*s*.

Oppenheim, A. N. *Questionnaire Design and Attitude Measurement*. London: Heinemann. 1966. Pp. ix + 298. 35*s*.

Osgood, C. E. *Perspective in Foreign Policy*. Palo Alto: Pacific Books. 1966. Pp. 94. Cloth $4.50. Paper $1.50.

Osler, S. F. & Cooke, R. E. *The Biosocial Basis of Mental Retardation*. Baltimore: Johns Hopkins. London: Oxford University Press. 1966. Pp. xiv + 151. 44*s*.

Ottaway, A. K. C. *Learning through Group Experience*. London: Routledge and Kegan Paul. 1966. Pp. viii + 168. 25*s*.

Palmer, J. O. & Goldstein, M. J. *Perspectives in Psychopathology*. New York: Oxford University Press. 1966. Pp. xiii + 434. 32*s*.

Pareek, U. N. *Developmental Patterns in Reactions to Frustration*. London: Asia Publishing House. 1965. Pp. viii + 182. 25*s*.

Parker, R. A. *Decision in Child Care*. London: Allen & Unwin. 1966. Pp. 120. 26*s*.

Petter, G. *Die geistige Entwicklung des Kindes im Werk von Jean Piaget*. Bern: Verlag. 1966. Pp. 367. Fr./DM 38.

Pepitone, A. *Attraction and Hostility*. London: Tavistock. 1966. Pp. xviii + 238. 30*s*.

Pikas, A. *Abstraction and Concept Formation*. Harvard: Harvard University Press. London: Oxford University Press. 1966. Pp. xiv + 303. 56*s*.

Post, F. *Persistent Persecutory States of the Elderly*. Oxford: Pergamon. 1966. Pp. ix + 100. 25*s*.

Rabkin, L. Y. *Psychopathology and Literature*. California: Chandler. 1966. xii + 326. No price given.

Ratcliffe, T. A. *The Development of Personality*. London: Allen and Unwin. 1967. Pp. 118. 20*s*.

Ramakrishna Rao, K. *Experimental Parapsychology*. Springfield, Ill.: Thomas. 1966. Pp. x + 255. $9.50.

Raychaudhuri, M. *Studies in Artistic Creativity*. Calcutta: Bharati. 1966. Pp. xii + 256. Rs. 15.00.

Rees, J. R. *Reflections. A Personal History and Account of the Growth of The World Federation for Mental Health*. New York: World Federation for Mental Health. 1966. Pp. xxviii + 192. No price given.

Reeves, J. W. *Thinking about Thinking*. London: Secker & Warburg. 1965. Pp. 333. 55*s*.

Reitman, W. R. *Cognition and Thought*. New York and London: Wiley. 1965. Pp. xii + 312. 60*s*

Ratterstol, N. *Paranoid and Paranoiac Psychoses*. Oslo: Universitetsforlaget. Pp. 259. N.Kr. 50.

Reuck, A. de & Knight, J. *Conflict in Society.* A Ciba Foundation Volume. London: Churchill. 1966. Pp. xii + 467. 65*s.*

Rey, A. *Les troubles de la mémoire et leur examen psychométrique.* Paris: Dessart. 1966. Pp. 192. F. 29.

Richardson, S. A., Dohrenwend, B. S. & Klein, D. *Interviewing: Its Forms and Functions.* London: Basic Books. 1965. Pp. viii + 380. 40*s.*

Robert, M. *The Psychoanalytic Revolution.* London: Allen and Unwin. 1966. Pp. 396. 50*s.*

Robinson, H. B. & Robinson, N. M. *The Mentally Retarded Child: a Psychological Approach.* New York: McGraw Hill. 1965. Pp. xvi + 639. 84*s.*

Robitscher, J. B. *Pursuit of Agreement: Psychiatry and the Law.* Philadelphia: Lippincott. 1966. Pp. 270. 70*s.*

Rokeach, M. *The Open and Closed Mind.* London: Basic Books. 1966. Pp. xv + 447. 27*s.* 6*d.*

Rorschach, H. *Gesammelte Autsaze.* Stuttgart: Verlag. 1965. Pp. 379. Fr. 38.

Rousey, C. L. & Moriarty, A. E. *Diagnostic Simplifications of Speech Sounds.* Springfield, Ill.: Thomas. 1966. Pp. xi + 156. $ 7.75.

Rycroft, C. *Psychoanalysis Observed.* London: Constable. 1966. Pp. 155. Cloth 21*s.* Paper 10*s.* 6*d.*

Rothwell, N. D. & Doniger, J. M. *The Psychiatric Half-way House.* Springfield, Ill.: Thomas. 1966. Pp. xvi + 265. $8.75.

Sarason, S. B. *et al. Psychology in Community Settings: Clinical, Educational, Vocational and Social Aspects.* London: Wiley. 1966. Pp. x + 714. 98*s.*

Sarnoff, I. *Society with Tears.* New York: Citadel Press. 1966. Pp. 318. $6.00.

Saugstad, P. *An Inquiry into the Foundations of Psychology.* London: Allen & Unwin. 1966. Pp. 104. 18*s.* 6*d.*

Scandinavian Studies in Criminology. Vol. I. Oslo: Universitets Forlaget. London: Tavistock. 1965. Pp. 226. 40*s.*

Schur, M. (ed.). *Drives, Affects, Behavior,* Vol. 2. New York: International Universities Press. 1965. Pp. 502. 90*s.*

Schwartz, M. S. & Shockley, E. L. *The Nurse and the Mental Patient.* New York: Wiley. 1966. Pp. 288. 14*s.*

Schmid, H. R. *Der Zullinger-test in der Berufsberatung.* Stuttgart: Verlag. 1965. Pp. 242. Fr./DM 28.

Sears, R. R., Rau, L. & Alpert, R. *Identification and Child Rearing.* London: Tavistock. 1966. Pp. xii + 382. 63*s.*

Semeonoff, B. *Personality Assessment.* Harmondsworth: Penguin. 1966. Pp. 443. 8*s.* 6*d.*

Serrano, J. R. *C. G. Jung and Hermann Hesse.* London: Routledge and Kegan Paul. 1966. Pp. ix + 112. 25*s.*

Shapiro, D. *Neurotic Styles.* London: Basic Books. 1966. Pp. xii + 208. 35*s.*

Sherif, M. & Cantril, H. *The Psychology of Ego-Involvements.* New York: Wiley. 1966. Pp. viii + 526.

Sidman, M. *Tactics of Scientific Research.* New York: Basic Books. 1966. Pp. x + 428. 27*s.* 6*d.*

Singer, J. L. *Daydreaming.* New York: Random House. 1966. Pp. xxi + 234. $2.25.

Slater, P. E. *Microcosm.* London: Wiley. 1966. Pp. xii + 276. 60*s.*

Slavson, S. R. *The Fields of Group Psychotherapy.* London and New York: Wiley. 1966. Pp. xii + 338. 15*s.*

Smythies, J. R. *Brain and Mind.* London: Routledge and Kegan Paul. 1965. Pp. x + 272. 40*s.*

Soviet Psychology and Psychiatry. Prepared for the XVIIIth International Congress of Psychology. Moscow 1966. Vol. IV. No. 3–4. New York: International Arts and Sciences Press. Pp. 146. 70*s.*

SPIELBERGER, C. D. *Anxiety and Behavior*. New York: Academic Press. 1966. Pp. xiv + 414. $9.75.

SPITZ, R. A. *The First Year of Life*. New York: International Universities Press. 1965. Pp. xix + 394. 77*s*.

SPROTT, W. J. H. *Social Psychology*. London: Social Sciences Paperbacks. 1966. Pp. xiv + 268. Cloth 35*s*.

SPROUT, H. & SPROUT, M. *The Ecological Perspective on Human Affairs*. Princeton: University Press. London: Oxford University Press. 1966. Pp. xi + 236. 44*s*.

STAFFORD-CLARK, D. *Psychiatry for Students*. London: Allen & Unwin. 1966. Pp. 276. 21*s*.

STAFFORD-CLARK, D. *What Freud Really Said*. London: MacDonald. 1966. Pp. 264. 18*s*.

STAGNER, R. & ROSEN, H. *The Psychology of Union-Management Relations*. London: Tavistock. 1966. Pp. vii + 147. 21.

STANILAND, A. C. *Patterns of Redundancy: A Psychological Study*. London: Cambridge University Press. 1966. Pp. viii + 216. 45*s*.

STEFAN, G. *In Search of Sanity: The Journal of a Schizophrenic*. New York: University Books. 1966. Pp. 224. $5.00.

STEPHENSON, G. *The Development of Conscience*. London: Routledge and Kegan Paul. 1966. Pp. x + 134. 25*s*.

STERNBACH, R. A. *Principles of Psychophysiology*. New York: Academic Press. 1966. Pp. xii + 297. 60*s*.

STEVENS, H. A. & HEBER, R. *Mental Retardation. A Review of Research*. Chicago and London: University of Chicago Press. Pp. xii + 502. 90*s*.

STOUFFER, S. A., GUTTMAN, L. & SUCHMAN, E. A. *Measurement and Prediction*. London: Wiley. 1966. Pp. x + 576. 23*s*.

STRAUS, E. W. *Phenomenological Psychology*. London: Tavistock. 1966. Pp. xiii + 353.

SUTHERLAND, R. L. & SMITH, B. K. *Understanding Mental Health*. Princeton: New Jersey: Van Nostrand. Pp. 234. 15*s*. 6*d*.

TAGIURI, R. & PETRULLO, L. (eds.). *Person Perception and Interpersonal Behavior*. Stanford: University Press and London: Oxford University Press. 1966. Pp. xvii + 390. 28*s*.

THE PSYCHO-ANALYTIC STUDY OF THE CHILD. Vol. 20. London: Hogarth. 1966. Pp. 566. 84*s*.

THINES, G. *Psychologie des animaux*. Brussels: Dessart. 1966. Pp. 352. No price given.

THOMAS, G. G. *et al*. *An Index of Responses to the Group Rorschach Test*. Baltimore: Johns Hopkins and London: Oxford University Press. Pp. xliv + 502. 120*s*.

THORPE, W. H. *Science, Man and Morals*. London: Methuen. 1965. Pp. xii + 176. 25*s*.

TIFFIN, J. & MCCORMICK, E. J. *Industrial Psychology*. London: Allen & Unwin. 1966. Pp. xiv + 682. 60*s*.

TORRANCE, E. P. & STROM, R. *Mental Health and Achievement*. London: Wiley. 1966. Pp. xiii + 417. 38*s*.

TOURNIER, P. (ed.). *Fatigue in Modern Society*. Richmond, Virginia: Knox. 1965. Pp. 79. $1.00.

UNDERWOOD, B. *Experimental Psychology*. New York: Appleton-Century-Crofts. 1966. Pp. ix + 678. $7.00.

USDIN, G. L. (ed.). *Psychoneurosis and Schizophrenia*. Philadelphia: Lippincott. 1966. Pp. 191. 60*s*.

VARNA, M. *An Introduction of Educational and Psychological Research*. London: Asia Publishing House. 1966. Pp. xi + 208. 40*s*.

VERNON, M. D. (ed.). *Experiments in Visual Perception*. Harmondsworth: Penguin. 1966. Pp. 443. 8*s*. 6*d*.

VERNON, M. D. *The Psychology of Perception*. London: University of London Press. 1965. Pp. 235. 25*s*.

C. V.

Volsky, T. *et al. The Outcomes of Counselling and Psychotherapy.* Minneapolis: University of Minnesota Press. London: Oxford University Press. 1965. Pp. xi + 209. 44*s.*

Von Haller Gilmer, B. *Industrial Psychology.* New York: McGraw Hill. 1966. Second Edition. Pp. xv + 617. 72*s.*

Waelder, R. *Psychoanalytic Avenues to Art.* London: Hogarth Press. 1965. Pp. 122. 25*s.*

Werkman, S. L. *The Role of Psychiatry in Medical Education.* London: Oxford University Press. 1966. Pp. xvi + 188. 38*s.*

Wertheimer, M. *Productive Thinking.* London: Tavistock. 1966. P. xvi + 302. Cloth 30*s.* Paper 21*s.*

Wessman, A. E. & Ricks, D. F. *Mood and Personality.* London: Holt, Rinehart and Winston. 1966. Pp. ix + 317. 48*s.*

Whipple, H. *Can Psychopathology be Measured?* New York: Annals of New York Academy of Sciences. 1965. Pp. 813–926. $4.00.

Whitaker, G. (ed.). *T-Group Training: Group Dynamics in Management Education.* ATM Occasional Papers 2. Oxford: Blackwell. 1965. Pp. v + 82. 14*s.*

White, A. R. *Attention.* Oxford: Blackwell. 1965. Pp. 131. 18*s.*

Wiener, N. & Schade, J. P. *Cybernetics of the Nervous System.* Progress in Brain Research. Vol. 17. Amsterdam: Elsevier. 1965. Pp. 424. No price given.

Wike, E. L. *Secondary Reinforcement: Selected Experiments.* New York: Harper and Row. 1966. Pp. xii + 516. No price given.

Williamson, R. C. *Marriage and Family Relations.* London: Wiley. 1966. Pp. *x* + 618. 60*s.*

Wittenborn, J. R. & Mary, P. R. *Prediction of Response to Pharmacotherapy.* Springfield, Ill.: Thomas. 1966. Pp. xii + 231. $10.00.

Wolberg, L. R. *Psychotherapy and the Behavioural Sciences.* London: Heinemann. 1966. Pp. 198. No price given.

Wortis, J. (ed.). *Recent Advances in Biological Psychiatry.* New York: Plenum Press. 1966. Pp. xii + 367. $16.50.

Wolpe, J. & Lazarus, A. A. *Behaviour Therapy Techniques.* Pergamon: Oxford. 1966. Pp. ix + 198.

Wyss, S. *Depth Psychology. A Critical Survey.* London: Allen and Unwin. 1966. Pp. 568. 52*s.*

Yates, A. (ed.). *Grouping in Education.* Hamburg: UNESCO Institute for Education and London: Wiley. 1966. Pp. 314. 96*s.*

Yates, A. J. *Frustration and Conflict.* London: Wiley. 1966. Pp. x + 236. 12*s.*

Yates, F. A. *The Art of Memory.* London: Routledge and Kegan Paul. 1966. Pp. xv + 400. 63*s.*

Zeppelin, I. S. V. *Die Variablen der Holtzman Inkblot Technique in ihrer Beziehung zur Introversion und Extraversion.* Stuttgart: Verlag. 1966. Pp. 102. Fr. 14.50.